Estimating the conditional extreme-value index under random right-censoring

Gilles Stupfler

Estimating the conditional extreme-value index under random right-censoring

Gilles Stupfler

Aix Marseille Université, CNRS, EHESS, Centrale Marseille, GREQAM UMR 7316,
13002 Marseille, France

Abstract. In extreme value theory, the extreme-value index is a parameter that controls the behavior of a cumulative distribution function in its right tail. Estimating this parameter is thus the first step when tackling a number of problems related to extreme events. In this paper, we introduce an estimator of the extreme-value index in the presence of a random covariate when the response variable is right-censored, whether its conditional distribution belongs to the Fréchet, Weibull or Gumbel domain of attraction. The pointwise weak consistency and asymptotic normality of the proposed estimator are established. Some illustrations on simulations are provided and we showcase the estimator on a real set of medical data.

AMS Subject Classifications: 62G05, 62G20, 62G30, 62G32, 62N01, 62N02.

Keywords: Extreme-value index, random covariate, random right-censoring, consistency, asymptotic normality.

1 Introduction

Studying extreme events is relevant in numerous fields of statistical applications. For instance, one can think about hydrology, where it is of interest to estimate the maximum level reached by seawater along a coast over a given period, or to study extreme rainfall at a given location; in actuarial science, a major problem for an insurance firm is to estimate the probability that a claim so large that it represents a threat to its solvency is filed. The focus in this type of problem is not in the estimation of "central" parameters of the random variable of interest, such as its mean or median, but rather in the understanding of its behavior in its right tail. The basic result in extreme value theory, known as the Fisher-Tippett-Gnedenko theorem (Fisher and Tippett [12], Gnedenko [16]) states that if (Y_n) is an independent sequence of random copies of a random variable Y such that there exist normalizing nonrandom sequences of real numbers (a_n) and (b_n), with $a_n > 0$ and such that the sequence

$$\frac{1}{a_n}\left(\max_{1\leq i\leq n} Y_i - b_n\right)$$

converges in distribution to some nondegenerate limit, then the cumulative distribution function (cdf) of this limit has the form $y \mapsto G_{\gamma_Y}(ay+b)$, with $a>0$ and b, $\gamma_Y \in \mathbb{R}$ where

$$G_{\gamma_Y}(y) = \begin{cases} \exp\left(-(1+\gamma_Y y)^{-1/\gamma_Y}\right) & \text{if } \gamma_Y \neq 0 \text{ and } 1+\gamma_Y y > 0, \\ \exp\left(-\exp(-y)\right) & \text{if } \gamma_Y = 0. \end{cases}$$

If the aforementioned convergence holds, we shall say that Y (or equivalently, its cdf F_Y) belongs to the domain of attraction (DA) of G_{γ_Y}, with γ_Y being the so-called extreme-value index of Y, and we write $F_Y \in \mathcal{D}(G_{\gamma_Y})$. The parameter γ_Y drives the behavior of G_{γ_Y} (and thus of F_Y) in its right tail:

- if $\gamma_Y > 0$, namely Y belongs to the Fréchet DA, then $1 - G_{\gamma_Y}$ is heavy-tailed *i.e.* it has a polynomial decay;
- if $\gamma_Y < 0$, namely Y belongs to the Weibull DA, then $1 - G_{\gamma_Y}$ is short-tailed *i.e.* it has a support bounded to the right;
- if $\gamma_Y = 0$, namely Y belongs to the Gumbel DA, then $1 - G_{\gamma_Y}$ has an exponential decay.

This makes it clear that the estimation of γ_Y is a first step when tackling various problems in extreme value analysis, such as the estimation of extreme quantiles of Y. Recent monographs on extreme value theory and especially univariate extreme-value index estimation include Beirlant *et al.* [2] and de Haan and Ferreira [19].

In practical applications, it may happen that only incomplete information is available. Consider for instance a medical follow-up study lasting up to time t which collects the survival times of patients for a given chronic disease. If a patient is diagnosed with the disease at time s, his/her survival time is known if and only if he/she dies before time t. If the patient survives until the end of the study, the only information available is that his/her survival time is not less than $t - s$. This situation is the archetypal example of right-censoring, which shall be the focus of this paper. An interesting problem in this particular case is the estimation of extreme survival times or, in other words, how long an exceptionally strong individual can survive the disease. A preliminary step necessary to give an answer to this question is to estimate the extreme-value index of the survival time Y; this problem, which is much more complex than the estimation of the extreme-value index when the data set is complete, has been investigated quite recently by Beirlant *et al.* [3] where asymptotic results for an extreme-value index estimator using the data above a nonrandom threshold are derived in the context of the Hall model (see Hall [20]), Einmahl *et al.* [11] in which the authors also suggest an estimator of extreme quantiles under random right-censoring so as to provide extreme survival times for male patients suffering from AIDS, Beirlant *et al.* [4] where maximum likelihood estimators are discussed, Sayah *et al.* [25] who focus on the heavy-tailed case and introduce a robust estimator with respect to contamination and Worms and Worms [29] where the consistency of several estimators, coming either from Kaplan-Meier integration or censored regression techniques, is studied. This situation should not be confused with right-truncation, in which case no information is available at all when Y is not actually observed: a recent reference in this case is Gardes and Stupfler [15].

Besides, it may well be the case that the survival time of a patient depends on additional random factors such as his/her age or the pre-existence of some other medical condition. Our goal in this study is to make it possible to integrate such information in the model by taking into account the dependency of Y on a covariate X. The problem thus becomes to estimate the conditional extreme-value index $\gamma_Y(x)$ of Y given $X = x$. Recent papers on this subject when Y is noncensored include Wang and Tsai [28] who introduced a maximum likelihood approach, Daouia *et al.* [7] who used a fixed number of nonparametric conditional quantile estimators to estimate the conditional extreme-value index, Gardes and Girard [13] who generalized the method of [7] to the case when the covariate space is infinite-dimensional, Goegebeur *et al.* [17] who studied a nonparametric regression estimator whose uniform asymptotic properties are examined in Goegebeur *et al.* [18] and Gardes and Stupfler [14] who introduced a smoothed local Hill estimator (see Hill [21]). All these papers consider the case when Y given $X = x$ belongs to the Fréchet DA; the case when the response distribution belongs to an arbitrary domain of attraction is considered in Daouia *et al.* [8], who generalized the method of [7] to this context and Stupfler [26] who introduced a generalization of the popular moment estimator of Dekkers *et al.* [10]. To the best of our knowledge, the only paper tackling this problem when Y is right-censored is Ndao *et al.* [23]; their work is, however, restricted to the case when Y is heavy-tailed. Our focus here is to devise an estimator which works regardless of whether or not the tail of Y is heavy.

The outline of this paper is as follows. In Section 2, we give a precise definition of our model. In Section 3, we define our estimator of the conditional extreme-value index. The pointwise weak consistency and asymptotic normality of the estimator are stated in Section 4. The finite sample performance of the estimator is studied in Section 5. In Section 6, we revisit the medical data set of [11] by integrating additional covariate information. Proofs are deferred to Section 7.

2 Framework

Let $(X_1, Y_1, C_1), \ldots, (X_n, Y_n, C_n)$ be n independent copies of a random vector (X, Y, C) taking its values in $E \times (0, \infty) \times (0, \infty)$ where E is a finite-dimensional linear space endowed with a norm $\| \cdot \|$. We assume that for all $x \in E$, given $X = x$, Y and C are independent, possess continuous probability density functions (pdfs) and that the related conditional survival functions (csfs) $\overline{F}_Y(\cdot|x) = 1 - F_Y(\cdot|x)$ of Y given $X = x$ and $\overline{F}_C(\cdot|x) = 1 - F_C(\cdot|x)$ of C given $X = x$ belong to some domain of attraction. Specifically, we shall work in the following setting, where we recall that the left-continuous inverse of a nondecreasing function f is the function $z \mapsto \inf\{y \in \mathbb{R} \mid f(y) \geq z\}$:

(M_1) Y and C are positive random variables and for every $x \in E$, there exist real numbers $\gamma_Y(x)$, $\gamma_C(x)$ and positive functions $a_Y(\cdot|x)$, $a_C(\cdot|x)$ such that the left-continuous inverses $U_Y(\cdot|x)$ of $1/\overline{F}_Y(\cdot|x)$ and $U_C(\cdot|x)$ of $1/\overline{F}_C(\cdot|x)$ satisfy

$$\lim_{t\to\infty} \frac{U_Y(tz|x) - U_Y(t|x)}{a_Y(t|x)} = D_{\gamma_Y(x)}(z) \quad \text{and} \quad \lim_{t\to\infty} \frac{U_C(tz|x) - U_C(t|x)}{a_C(t|x)} = D_{\gamma_C(x)}(z)$$

for every $z > 0$, where

$$D_\gamma(z) = \begin{cases} \dfrac{z^\gamma - 1}{\gamma} & \text{if } \gamma \neq 0 \\ \log z & \text{if } \gamma = 0. \end{cases} \tag{1}$$

Model (M_1) is the conditional analogue of the classical extreme-value framework for Y and C, see for instance [19], p.19. In this model, for every $x \in E$, the functions $U_Y(\cdot|x)$ and $U_C(\cdot|x)$ have positive limits $U_Y(\infty|x)$ and $U_C(\infty|x)$ at infinity; the functions $x \mapsto U_Y(\infty|x)$ and $x \mapsto U_C(\infty|x)$, which are such that

$$U_Y(\infty|x) = \sup\{t \in \mathbb{R} \,|\, \overline{F}_Y(t|x) > 0\} \text{ and } U_C(\infty|x) = \sup\{t \in \mathbb{R} \,|\, \overline{F}_C(t|x) > 0\}$$

are respectively called the conditional right endpoints of Y and C.

We assume that we only observe the random vectors (X_i, T_i, δ_i) with $T_i = Y_i \wedge C_i$ and $\delta_i = \mathbb{I}_{\{Y_i \leq C_i\}}$, where we denote by $s \wedge t$ the minimum of s and t. Suppose that the following condition holds as well:

($\mathcal{H}$) For every $x \in E$, the distribution of T given $X = x$ belongs to some domain of attraction $\mathcal{D}(G_{\gamma_T(x)})$ and we have either

- $\gamma_Y(x) > 0$ and $\gamma_C(x) > 0$;
- $\gamma_Y(x) < 0$, $\gamma_C(x) < 0$ and $0 < U_Y(\infty|x) = U_C(\infty|x) < \infty$;
- $\gamma_Y(x) = \gamma_C(x) = 0$ and $U_Y(\infty|x) = U_C(\infty|x) = \infty$.

An unconditional analogue of hypothesis ($\mathcal{H}$) is condition (7) in [11]. This hypothesis ensures that one works in an interesting case regarding censoring: censoring in the extremes of the sample should be present to justify using an adapted methodology, but not complete so that we can expect to recover information about the extremes of Y. For instance, any situation in which the tail of Y is heavy and C is short-tailed, namely $\gamma_Y(x) > 0$ and $\gamma_C(x) < 0$, is a so-called "completely censored situation" in the extremes (see also [11]): clearly, high values of Y exceeding the right endpoint of C will be censored with probability 1 and this makes it impossible to recover anything about the right tail of Y. If on the contrary the random variable Y is short-tailed and the tail of C is heavy, corresponding to the case $\gamma_Y(x) < 0$ and $\gamma_C(x) > 0$, then we are in an "uncensored situation" in which high values of T come from high values of Y with probability approaching 1 as the sample size increases. In this situation, a methodology such as the one of [26] which does not account for the right-censoring phenomenon will yield essentially the same results as an adapted technique provided the sample size is large enough.

As mentioned in [11], if (M_1) and ($\mathcal{H}$) hold then T has conditional right endpoint $U_Y(\infty|x) = U_C(\infty|x)$ and conditional extreme-value index

$$\gamma_T(x) = \frac{\gamma_Y(x)\gamma_C(x)}{\gamma_Y(x) + \gamma_C(x)}$$

with the convention $\gamma_T(x) = 0$ if $\gamma_Y(x) = \gamma_C(x) = 0$. In other words, if $\overline{F}_T(\cdot|x)$ is the csf of T given $X = x$, there exists a positive function $a_T(\cdot|x)$ such that the left-continuous inverse $U_T(\cdot|x)$ of $1/\overline{F}_T(\cdot|x)$ satisfies

$$\forall z > 0, \lim_{t \to \infty} \frac{U_T(tz|x) - U_T(t|x)}{a_T(t|x)} = D_{\gamma_T(x)}(z).$$

3 The estimators

To tackle the problem, we start by introducing an estimator of the conditional extreme-value index γ_T. For $x \in E$ and a sequence $h = h(n)$ converging to 0 as $n \to \infty$, we let $N_n(x, h)$ be the total number of observations in the closed ball $B(x, h)$ with center x and radius h:

$$N_n(x,h) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i \in B(x,h)\}} \text{ with } B(x,h) = \{x' \in E \,|\, \|x - x'\| \leq h\},$$

where $\mathbb{I}_{\{.\}}$ is the indicator function. The bandwidth sequence $h(n)$ makes it possible to select those covariates which are close enough to x. Given $N_n(x,h) = l \geq 1$, we let, for $i = 1, \ldots, l$, $(\mathcal{T}_i, \Delta_i) = (\mathcal{T}_i(x,h), \Delta_i(x,h))$ be the response pairs whose associated covariates $\mathcal{X}_i = \mathcal{X}_i(x,h)$ belong to the ball $B(x,h)$. Let $\mathcal{T}_{1,l} \leq \cdots \leq \mathcal{T}_{l,l}$ be the order statistics associated with the sample $(\mathcal{T}_1, \ldots, \mathcal{T}_l)$ – this way of denoting order statistics shall be used throughout the paper – and set for $j = 1, 2$:

$$M_n^{(j)}(x, k_x, h) = \frac{1}{k_x} \sum_{i=1}^{k_x} [\log(\mathcal{T}_{l-i+1,l}) - \log(\mathcal{T}_{l-k_x,l})]^j$$

if $k_x \in \{1, \ldots, l-1\}$ and 0 otherwise. Define:

$$\begin{aligned} \widehat{\gamma}_{T,n}(x,k_x,h) &= \widehat{\gamma}_{T,n,+}(x,k_x,h) + \widehat{\gamma}_{T,n,-}(x,k_x,h) \\ \text{where } \widehat{\gamma}_{T,n,+}(x,k_x,h) &= M_n^{(1)}(x,k_x,h) \\ \text{and } \widehat{\gamma}_{T,n,-}(x,k_x,h) &= 1 - \frac{1}{2}\left(1 - \frac{\left[M_n^{(1)}(x,k_x,h)\right]^2}{M_n^{(2)}(x,k_x,h)}\right)^{-1} \end{aligned}$$

if $\left[M_n^{(1)}(x,k_x,h)\right]^2 \neq M_n^{(2)}(x,k_x,h)$, with $\widehat{\gamma}_{T,n,-}(x,k_x,h) = 0$ otherwise. The estimator $\widehat{\gamma}_{T,n}(x,k_x,h)$ is an adaptation of the moment estimator of [10] to the presence of a random covariate; it follows from Theorem 1 in [26] that this quantity is a pointwise consistent estimator of the extreme-value index $\gamma_T(x)$ of T given $X = x$ under mild conditions.

We then adapt an idea of [11] in order to obtain an estimator of $\gamma_Y(x)$. Given $N_n(x,h) = l$, let $\Delta_{[1:l]}, \ldots, \Delta_{[l:l]}$ be the order statistics induced by the sample $(\mathcal{T}_1, \ldots, \mathcal{T}_l)$: $\Delta_{[i:l]}$ is the random variable associated with $\mathcal{T}_{i,l}$. We define

$$\widehat{p}_n(x,k_x,h) = \frac{1}{k_x} \sum_{i=1}^{k_x} \Delta_{[l-i+1:l]},$$

the proportion of noncensored observations among $\mathcal{T}_{l-k_x+1,l}, \ldots, \mathcal{T}_{l,l}$ when $k_x \in \{1, \ldots, l-1\}$ and 0 otherwise. This estimator is the adaptation to the random covariate case of the estimator $\widehat{p}$ of [11]: we shall show (see the proof of Theorem 1 below) that under some conditions, $\widehat{p}_n(x,k_x,h)$ is a consistent estimator of $\gamma_C(x)/(\gamma_Y(x) + \gamma_C(x))$. Our estimator of $\gamma_Y(x)$ is then

$$\widehat{\gamma}_{Y,n}(x,k_x,h) = \frac{\widehat{\gamma}_{T,n}(x,k_x,h)}{\widehat{p}_n(x,k_x,h)}$$

if $\widehat{p}_n(x,k_x,h) > 0$ and 0 otherwise.

4 Main results

4.1 Weak consistency

We start by giving a pointwise weak consistency result for our estimator. To this end let $n_x = n_x(n,h) = n\mathbb{P}(X \in B(x,h))$ be the average total number of points in the ball $B(x,h)$ and assume that $n_x(n,h) > 0$ for every n. Let $k_x = k_x(n)$ be a sequence of positive integers, $F_{T,h}(\cdot|x)$ be the cdf of T given $X \in B(x,h)$ and $U_{T,h}(\cdot|x)$ be the left-continuous inverse of $1/\overline{F}_{T,h}(\cdot|x)$. We introduce the functions $p(\cdot|x)$, $p_h(\cdot|x)$ defined by

$$\begin{aligned} p(t|x) &= \frac{d}{dt}\mathbb{P}(\delta = 1, T \leq t \,|\, X = x) \Big/ \frac{d}{dt} F_T(t|x) \\ \text{and} \ \ p_h(t|x) &= \frac{d}{dt}\mathbb{P}(\delta = 1, T \leq t \,|\, X \in B(x,h)) \Big/ \frac{d}{dt} F_{T,h}(t|x) \end{aligned}$$

for every $t > 0$ such that the denominator is nonzero and $p(x) := \gamma_C(x)/(\gamma_Y(x) + \gamma_C(x))$ otherwise. It follows from Lemma 1 (see Section 7) that if (M_1), $(\mathcal{H})$ hold and $\gamma_Y(x) \neq 0$, then the first of these two quantities converges to the positive limit $p(x)$ as $t \to U_T(\infty|x)$ and from Lemma 2 that the second quantity is indeed well-defined. Assume that in the case $\gamma_Y(x) = \gamma_C(x) = 0$, the function $p(\cdot|x)$ also converges to a positive limit at infinity, which we denote by $p(x)$ for the sake of consistency. The function $x \mapsto 1 - p(x)$ is understood as the conditional percentage of censoring in the right tail of Y. For u, $v \in (1,\infty)$ such that $u < v$, we introduce the quantities

$$\begin{aligned} \omega(\log U_T, u, v, x, h) &= \sup_{t \in [u,v]} \left| \log \frac{U_{T,h}(t|x)}{U_T(t|x)} \right| \\ \text{and} \ \ \omega(p \circ U_T, u, v, x, h) &= \sup_{t \in [u,v]} \left| p_h(U_{T,h}(t|x)|x) - p(x) \right|. \end{aligned}$$

Our consistency result is then:

Theorem 1. *Assume that (M_1) and $(\mathcal{H})$ hold. For some $x \in E$, assume that $n_x \to \infty$, $k_x \to \infty$, $k_x/n_x \to 0$ and for some $\eta > 0$*

$$\frac{U_T(n_x/k_x|x)}{a_T(n_x/k_x|x)} \omega\left(\log U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) \to 0 \ \ \text{as} \ \ n \to \infty \tag{2}$$

$$\text{and} \ \ \omega\left(p \circ U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) \to 0 \ \ \text{as} \ \ n \to \infty. \tag{3}$$

Then it holds that $\widehat{\gamma}_{Y,n}(x, k_x, h) \xrightarrow{\mathbb{P}} \gamma_Y(x)$ as $n \to \infty$.

Conditions $k_x \to \infty$ and $k_x/n_x \to 0$ in Theorem 1 are standard hypotheses for the estimation of the conditional extreme-value index. Moreover, condition $n_x \to \infty$ is necessary to make sure that there are sufficiently many observations close to x, which is a standard assumption in the random covariate case.

We conclude this section by analyzing conditions (2) and (3). We assume that

(A_1) For every $x \in E$, it holds that for all $t \in (0, U_T(\infty|x))$, $f_Y(t|x) > 0$ and $f_C(t|x) > 0$.

(A_2) The functions γ_Y and γ_C are continuous functions on E.

(A_3) For every $x' \in B(x,h)$ and $r > 0$, we have $\mathbb{P}(X \in B(x',r)) > 0$ if n is large enough.

(A_4) For every $y \in \mathbb{R}$, the function $\overline{F}_T(y|\cdot)$ is continuous on E.

Note that hypothesis (A_1) implies that the csf $\overline{F}_T(\cdot|x)$ is a continuous decreasing function on $(0, U_T(\infty|x))$ and hypothesis (A_2) entails that the function γ_T is continuous. Hypotheses (A_3) and (A_4) are technical conditions; see Proposition 1 in [26] for analogues of these assumptions in the noncensored case. We can draw two consequences from this:

1. If $\gamma_Y(x) > 0$ and $\gamma_C(x) > 0$ then $\gamma_Y(x') > 0$, $\gamma_C(x') > 0$ and $\gamma_T(x') > 0$ for x' close enough to x. Corollary 1.2.10 in [19], p.23 thus yields for n large enough and every $x' \in B(x,h)$
$$\forall z > 1,\ U_T(z|x') = z^{\gamma_T(x')} L_{U_T}(z|x')$$
where for every $x' \in B(x,h)$, $L_{U_T}(\cdot|x')$ is a slowly varying function at infinity, and
$$\forall t > 0,\ \overline{F}_Y(t|x') = t^{-1/\gamma_Y(x')} L_{\overline{F}_Y}(t|x') \ \text{ and } \ \overline{F}_C(t|x') = t^{-1/\gamma_C(x')} L_{\overline{F}_C}(t|x')$$
where $L_{\overline{F}_Y}(\cdot|x')$ and $L_{\overline{F}_C}(\cdot|x')$ are continuously differentiable slowly varying functions at infinity. Especially, if
$$b_Y(t|x') = t\frac{L'_{\overline{F}_Y}(t|x')}{L_{\overline{F}_Y}(t|x')} \ \text{ and } \ b_C(t|x') = t\frac{L'_{\overline{F}_C}(t|x')}{L_{\overline{F}_C}(t|x')}$$
then
$$\begin{aligned} \forall t > 0,\ f_Y(t|x') &= \left[\frac{1}{\gamma_Y(x')} - b_Y(t|x')\right] \frac{\overline{F}_Y(t|x')}{t} \\ \text{and } \ f_C(t|x') &= \left[\frac{1}{\gamma_C(x')} - b_C(t|x')\right] \frac{\overline{F}_C(t|x')}{t}. \end{aligned}$$

2. If $\gamma_Y(x) < 0$ and $\gamma_C(x) < 0$ then $\gamma_Y(x') < 0$, $\gamma_C(x') < 0$ and $\gamma_T(x') < 0$ for x' close enough to x. Corollary 1.2.10 in [19], p.23 yields for n large enough and every $x' \in B(x,h)$ that
$$\forall z > 1,\ U_T(\infty|x') - U_T(z|x') = z^{\gamma_T(x')} L_{U_T}(z|x')$$
where for every $x' \in B(x,h)$, $L_{U_T}(\cdot|x')$ is a slowly varying function at infinity and
$$\begin{aligned} \forall t > 0,\ \overline{F}_Y(U_Y(\infty|x') - t^{-1}|x') &= t^{1/\gamma_Y(x')} L_{\overline{F}_Y}(t|x') \\ \text{and } \ \overline{F}_C(U_C(\infty|x') - t^{-1}|x') &= t^{1/\gamma_C(x')} L_{\overline{F}_C}(t|x') \end{aligned}$$
where $L_{\overline{F}_Y}(\cdot|x')$ and $L_{\overline{F}_C}(\cdot|x')$ are continuously differentiable slowly varying functions at infinity. In particular, if
$$\begin{aligned} b_Y(t|x') &= \begin{cases} t\dfrac{L'_{\overline{F}_Y}(t|x')}{L_{\overline{F}_Y}(t|x')} & \text{if } \ L_{\overline{F}_Y}(t|x') > 0 \\ 0 & \text{otherwise} \end{cases} \\ \text{and } \ b_C(t|x') &= \begin{cases} t\dfrac{L'_{\overline{F}_C}(t|x')}{L_{\overline{F}_C}(t|x')} & \text{if } \ L_{\overline{F}_C}(t|x') > 0 \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

then, recalling that $U_T(\infty|x') = U_Y(\infty|x') = U_C(\infty|x')$ for every $x' \in B(x,h)$, one may write

$$\begin{aligned}
\forall t \in (0, U_T(\infty|x')),\ f_Y(t|x') &= \left[-\frac{1}{\gamma_Y(x')} - b_Y((U_T(\infty|x')-t)^{-1}|x')\right] \frac{\overline{F}_Y(t|x')}{U_T(\infty|x')-t} \\
\text{and}\ \ f_C(t|x') &= \left[-\frac{1}{\gamma_C(x')} - b_C((U_T(\infty|x')-t)^{-1}|x')\right] \frac{\overline{F}_C(t|x')}{U_T(\infty|x')-t}.
\end{aligned}$$

In this framework, it is possible to reformulate the hypotheses in our main results in a more convenient fashion: let $K_{x,\eta} := [n_x/(1+\eta)k_x, n_x^{1+\eta}]$ and assume that for some $\alpha \in (0,1]$

$$\begin{aligned}
\sup_{x' \in B(x,h)} |\gamma_Y(x') - \gamma_Y(x)| \vee |\gamma_C(x') - \gamma_C(x)| &= \mathrm{O}\left(h^\alpha\right) \quad (4) \\
\text{and} \quad \sup_{z \in K_{x,\eta}} \sup_{x' \in B(x,h)} \frac{1}{\log z} \left|\log \frac{L_{U_T}(z|x')}{L_{U_T}(z|x)}\right| &= \mathrm{O}\left(h^\alpha\right) \quad (5)
\end{aligned}$$

where we denote by $s \vee t$ the maximum of two real numbers s and t. Then in case 1, if $h^\alpha \log n_x \to 0$ as $n \to \infty$, one has

$$\frac{U_T(n_x/k_x|x)}{a_T(n_x/k_x|x)} \omega\left(\log U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) = \mathrm{O}(h^\alpha \log n_x) \quad (6)$$

see the discussion below Proposition 1 in [26]. In case 2, if the conditional right endpoint $U_T(\infty|\cdot)$ is such that

$$\sup_{x' \in B(x,h)} |U_T(\infty|x') - U_T(\infty|x)| = \mathrm{O}(h^\beta) \quad (7)$$

with $\beta \in (0,1]$, then if

$$h^\alpha \log n_x \to 0 \ \text{ and } \ \frac{(n_x/k_x)^{-\gamma_T(x)}}{L_{U_T}(n_x/k_x|x)} h^\beta \to 0 \ \text{ as } \ n \to \infty \quad (8)$$

one has

$$\frac{U_T(n_x/k_x|x)}{a_T(n_x/k_x|x)} \omega\left(\log U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) = \mathrm{O}\left(h^\alpha \log n_x \vee \frac{(n_x/k_x)^{-\gamma_T(x)}}{L_{U_T}(n_x/k_x|x)} h^\beta\right), \quad (9)$$

see again the discussion below Proposition 1 in [26].

The next result gives bounds of this kind when considering hypothesis (3).

Proposition 1. *Assume that conditions (M_1), $(\mathcal{H})$, (A_1), (A_2), (A_3), (A_4) hold and that for some $\alpha \in (0,1]$ and $\eta > 0$, conditions (4) and (5) are satisfied.*

1. *In case 1 above assume that for x' close enough to x, $|b_Y(\cdot|x')|$, $|b_C(\cdot|x')|$ are regularly varying functions at infinity with respective indices $\rho_Y(x')/\gamma_Y(x')$, $\rho_C(x')/\gamma_C(x')$, that is*

$$|b_Y(t|x')| = t^{\rho_Y(x')/\gamma_Y(x')} L_{b_Y}(t|x') \quad \text{and} \quad |b_C(t|x')| = t^{\rho_C(x')/\gamma_C(x')} L_{b_C}(t|x') \quad (10)$$

with the so-called conditional second-order parameter functions ρ_Y, ρ_C and the slowly varying functions $L_{b_Y}(\cdot|x')$, $L_{b_C}(\cdot|x')$ satisfying, for some $\eta > 0$,

$$\sup_{x' \in B(x,h)} |\rho_Y(x') - \rho_Y(x)| \vee |\rho_C(x') - \rho_C(x)| \quad = \quad \mathrm{O}\left(h^\alpha\right), \tag{11}$$

$$\sup_{t \in U_T(K_{x,\eta}|x)} \sup_{x' \in B(x,h)} \frac{1}{\log t} \left[\left| \log \frac{L_{b_Y}(t|x')}{L_{b_Y}(t|x)} \right| \vee \left| \log \frac{L_{b_C}(t|x')}{L_{b_C}(t|x)} \right| \right] \quad = \quad \mathrm{O}\left(h^\alpha\right) \tag{12}$$

where $U_T(K_{x,\eta}|x)$ is the image of the interval $K_{x,\eta}$ by the function $U_T(\cdot|x)$. If $\rho_Y(x)$ and $\rho_C(x)$ are negative, $h^\alpha \log n_x \to 0$ and the sequence

$$\delta_n := |b_Y(U_T(n_x/k_x|x)|x)| \vee |b_C(U_T(n_x/k_x|x)|x)|$$

converges to 0 then, for $\eta > 0$ small enough, one has, as $n \to \infty$:

$$\omega\left(p \circ U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) = \mathrm{O}\left(h^\alpha \log n_x \vee \delta_n\right).$$

2. *In case 2 above assume that conditions (7) and (8) are satisfied. Assume moreover that for x' close enough to x, $|b_Y(\cdot|x')|$ and $|b_C(\cdot|x')|$ are regularly varying functions at infinity with respective indices $-\rho_Y(x')/\gamma_Y(x')$ and $-\rho_C(x')/\gamma_C(x')$, namely*

$$|b_Y(t|x')| = t^{-\rho_Y(x')/\gamma_Y(x')} L_{b_Y}(t|x') \quad \text{and} \quad |b_C(t|x')| = t^{-\rho_C(x')/\gamma_C(x')} L_{b_C}(t|x') \tag{13}$$

with the conditional second-order parameter functions ρ_Y, ρ_C satisfying (11) and the slowly varying functions $L_{b_Y}(\cdot|x')$, $L_{b_C}(\cdot|x')$ being such that for some $\eta \in (0,1)$

$$\sup_{t \in J_{x,\eta}} \sup_{x' \in B(x,h)} \frac{1}{\log t} \left[\left| \log \frac{L_{b_Y}(t|x')}{L_{b_Y}(t|x)} \right| \vee \left| \log \frac{L_{b_C}(t|x')}{L_{b_C}(t|x)} \right| \right] = \mathrm{O}\left(h^\alpha\right) \tag{14}$$

where $J_{x,\eta} := [(1-\eta)[U_T(\infty|x) - U_T(n_x/k_x|x)]^{-1}, \infty)$. If $\rho_Y(x)$ and $\rho_C(x)$ are negative and the sequence

$$\delta_n := |b_Y((U_T(\infty|x) - U_T(n_x/k_x|x))^{-1}|x)| \vee |b_C((U_T(\infty|x) - U_T(n_x/k_x|x))^{-1}|x)|$$

converges to 0 then one has, as $n \to \infty$:

$$\omega\left(p \circ U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) = \mathrm{O}\left(h^\alpha \log n_x \vee \delta_n\right).$$

This result relates hypothesis (3) in Theorem 1 to the various functions involved in the usual parametrization of the problem. It shall allow us to recover the optimal rate of convergence of the estimator, see Theorem 2 and the developments below for details.

4.2 Asymptotic normality

To prove a pointwise asymptotic normality result for our estimator, we need to introduce a second-order condition on the function $U_T(\cdot|x)$:

(M_2) Conditions (M_1) and $(\mathcal{H})$ hold and for every $x \in E$, there exist a real number $\rho_T(x) \leq 0$ and a function $A_T(\cdot|x)$ of constant sign converging to 0 at infinity such that the function $U_T(\cdot|x)$ satisfies

$$\forall z > 0, \ \lim_{t \to \infty} \frac{1}{A_T(t|x)} \left[\frac{U_T(tz|x) - U_T(t|x)}{a_T(t|x)} - D_{\gamma_T(x)}(z) \right] = H_{\gamma_T(x), \rho_T(x)}(z)$$

where

$$H_{\gamma_T(x), \rho_T(x)}(z) = \int_1^z r^{\gamma_T(x)-1} \left[\int_1^r s^{\rho_T(x)-1} ds \right] dr.$$

Hypothesis (M_2) is the conditional analogue of a classical second-order condition, see for instance Definition 2.3.1 and Corollary 2.3.4 in [19], pp.44–45: the parameter $\rho_T(x)$ is the so-called second-order parameter of T given $X = x$. Note that Theorem 2.3.3 in [19], p.44 shows that the function $|A_T(\cdot|x)|$ is regularly varying at infinity with index $\rho_T(x)$. Moreover, as shown in Lemma B.3.16 p.397 therein, if (M_2) holds with $\gamma_T(x) \neq \rho_T(x)$ and $\rho_T(x) < 0$ if $\gamma_T(x) > 0$, then defining $q_T(\cdot|x) = a_T(\cdot|x)/U_T(\cdot|x)$, a second-order condition also holds for the function $\log U_T(\cdot|x)$, namely:

$$\forall z > 0, \ \lim_{t \to \infty} \frac{1}{Q_T(t|x)} \left[\frac{\log U_T(tz|x) - \log U_T(t|x)}{q_T(t|x)} - D_{\gamma_{T,-}(x)}(z) \right] = H_{\gamma_{T,-}(x), \rho'_T(x)}(z)$$

with $\gamma_{T,-}(x) = \gamma_T(x) \wedge 0$,

$$\rho'_T(x) = \begin{cases} \rho_T(x) & \text{if } \gamma_T(x) < \rho_T(x) \leq 0 \\ \gamma_T(x) & \text{if } \rho_T(x) < \gamma_T(x) \leq 0 \\ -\gamma_T(x) & \text{if } 0 < \gamma_T(x) < -\rho_T(x) \text{ and } \ell_T(x) \neq 0 \\ \rho_T(x) & \text{if } (0 < \gamma_T(x) < -\rho_T(x) \text{ and } \ell_T(x) = 0) \text{ or } 0 < -\rho_T(x) \leq \gamma_T(x) \end{cases}$$

where we have defined

$$\ell_T(x) = \lim_{t \to \infty} \left(U_T(t|x) - \frac{a_T(t|x)}{\gamma_T(x)} \right)$$

and $Q_T(\cdot|x)$ has ultimately constant sign, converges to 0 at infinity and is such that $|Q_T(\cdot|x)|$ is regularly varying at infinity with index $\rho'_T(x)$. Note that Lemma B.3.16 in [19], p.397 entails that one can choose

$$Q_T(t|x) = \begin{cases} A_T(t|x) & \text{if} \quad \gamma_T(x) < \rho_T(x) \leq 0 \\[2ex] \gamma_{T,+}(x) - \dfrac{a_T(t|x)}{U_T(t|x)} & \begin{array}{ll} \text{if} & \rho_T(x) < \gamma_T(x) \leq 0 \\ \text{or} & 0 < \gamma_T(x) < -\rho_T(x) \text{ and } \ell_T(x) \neq 0 \\ \text{or} & 0 < \gamma_T(x) = -\rho_T(x) \end{array} \\[4ex] \dfrac{\rho_T(x)}{\gamma_T(x) + \rho_T(x)} A_T(t|x) & \begin{array}{ll} \text{if} & 0 < \gamma_T(x) < -\rho_T(x) \text{ and } \ell_T(x) = 0 \\ \text{or} & 0 < -\rho_T(x) < \gamma_T(x) \end{array} \end{cases}$$

with $\gamma_{T,+}(x) = \gamma_T(x) \vee 0$. Besides, if $\gamma_T(x) > 0$ and $\rho_T(x) = 0$, then one has

$$\forall z > 0, \ \lim_{t \to \infty} \frac{1}{Q_T(t|x)} \left[\frac{\log U_T(tz|x) - \log U_T(t|x)}{q_T(t|x)} - \log z \right] = 0$$

for every function $Q_T(\cdot|x)$ such that $A_T(t|x) = \mathrm{O}(Q_T(t|x))$ as $t \to \infty$; especially, we can and will take $Q_T(\cdot|x) = A_T(\cdot|x)$ and $\rho'_T(x) = 0$ in this case.

We can now state the asymptotic normality of our estimator.

Theorem 2. *Assume that (M_2) holds. For some $x \in E$, assume that $n_x \to \infty$, $k_x \to \infty$, $k_x/n_x \to 0$, $\sqrt{k_x}\, Q_T(n_x/k_x|x) \to 0$ and for some $\eta > 0$*

$$\sqrt{k_x}\frac{U_T(n_x/k_x|x)}{a_T(n_x/k_x|x)}\omega\left(\log U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) \quad \to \quad 0 \quad \textit{as} \quad n \to \infty \tag{15}$$

$$\textit{and} \quad \sqrt{k_x}\,\omega\left(p \circ U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) \quad \to \quad 0 \quad \textit{as} \quad n \to \infty. \tag{16}$$

Then if $\gamma_T(x) \neq \rho_T(x)$, it holds that

$$\sqrt{k_x}\,[\widehat{\gamma}_{Y,n}(x, k_x, h) - \gamma_Y(x)] \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{p^2(x)}\left[V(\gamma_T(x)) + \frac{\gamma_T^2(x)}{p(x)}(1 - p(x))\right]\right)$$

where we have set

$$V(\gamma_T(x)) = \begin{cases} \gamma_T^2(x) + 1 & \textit{if } \gamma_T(x) \geq 0 \\ \dfrac{(1-\gamma_T(x))^2(1-2\gamma_T(x))(1-\gamma_T(x)+6\gamma_T^2(x))}{(1-3\gamma_T(x))(1-4\gamma_T(x))} & \textit{if } \gamma_T(x) < 0. \end{cases}$$

Theorem 2 is the conditional analogue of the asymptotic normality result stated in [11]. In particular, the asymptotic variance of our estimator is similar to the one obtained when there is no covariate. Besides, condition $\sqrt{k_x}\, Q_T(n_x/k_x|x) \to 0$ as $n \to \infty$ in Theorem 2 is a standard condition needed to control the bias of the estimator.

We conclude this paragraph by showing how Theorem 2 can be used to obtain optimal rates of convergence for our estimator. We assume that $E = \mathbb{R}^d$, $d \geq 1$ is equipped with the standard Euclidean distance and that X has a probability density function f on $\mathbb{R}^d$ which is continuous on its support S, assumed to have nonempty interior. If x is a point lying in the interior of S which is such that $f(x) > 0$, it is straightforward to show that (A_3) holds and that

$$n_x = n\int_{B(x,h)} f(u)du = nh^d\mathcal{V}f(x)(1 + \mathrm{o}(1)) \quad \text{as} \quad n \to \infty$$

with $\mathcal{V}$ being the volume of the unit ball in $\mathbb{R}^d$. Set $k = k_x/(h^d\mathcal{V}f(x))$; it is then clear that $k_x = kh^d\mathcal{V}f(x)$ and that hypotheses $n_x \to \infty$, $k_x \to \infty$ and $k_x/n_x \to 0$ as $n \to \infty$ are equivalent to $kh^d \to \infty$ and $k/n \to 0$ as $n \to \infty$. If k and h have respective order n^a and n^{-b}, with a, $b > 0$, the rate of convergence of the estimator $\widehat{\gamma}_{Y,n}(x, k_x, h)$ to $\gamma_Y(x)$ is then $n^{(a-bd)/2}$. Under the hypotheses of Theorem 2, provided that (A_1), (A_2) and (A_4) hold, one can find the optimal values for a and b in the Fréchet and Weibull domains of attraction:

- If $\gamma_Y(x) > 0$ and $\gamma_C(x) > 0$, then under the Hölder conditions (4) and (5), hypothesis (15) shall be satisfied if $\sqrt{kh^d}h^\alpha \log(nh^d) \to 0$ as $n \to \infty$. Besides, under assumption (10) and the Hölder conditions (11) and (12), Proposition 1 gives that hypothesis (16) is implied by

$$\sqrt{k_x}\left[|b_Y(U_T(n_x/k_x|x)|x)| \vee |b_C(U_T(n_x/k_x|x)|x)|\right] \to 0 \quad \text{as} \quad n \to \infty$$

or, equivalently,

$$\begin{aligned}\sqrt{kh^d}\left(\frac{n}{k}\right)^{\gamma_C(x)\rho_Y(x)/(\gamma_Y(x)+\gamma_C(x))}\mathcal{L}_Y(n/k|x) &\to 0\\ \text{and}\quad \sqrt{kh^d}\left(\frac{n}{k}\right)^{\gamma_Y(x)\rho_C(x)/(\gamma_Y(x)+\gamma_C(x))}\mathcal{L}_C(n/k|x) &\to 0\end{aligned}$$

as $n\to\infty$, where $\mathcal{L}_C(\cdot|x)$ and $\mathcal{L}_Y(\cdot|x)$ are slowly varying functions at infinity. Recalling the bias condition $\sqrt{kh^d}\,Q_T(n/k|x)\to 0$ as $n\to\infty$, the problem is thus to maximize the quantity $a-bd$ under the constraints $a\in(0,1)$, $a-bd\geq 0$,

$$\begin{aligned}a-b(d+2\alpha) &\leq 0,\\ a-bd+2(1-a)\frac{\gamma_C(x)\rho_Y(x)}{\gamma_Y(x)+\gamma_C(x)} &\leq 0,\\ a-bd+2(1-a)\frac{\gamma_Y(x)\rho_C(x)}{\gamma_Y(x)+\gamma_C(x)} &\leq 0\\ \text{and}\quad a-bd+2(1-a)\rho_T'(x) &\leq 0.\end{aligned}$$

Setting

$$\rho(x):=\max\left(\rho_T'(x),\frac{\gamma_C(x)\rho_Y(x)}{\gamma_Y(x)+\gamma_C(x)},\frac{\gamma_Y(x)\rho_C(x)}{\gamma_Y(x)+\gamma_C(x)}\right)\leq 0$$

the constraints become $a\in(0,1)$, $a-bd\geq 0$,

$$a-b(d+2\alpha)\leq 0\quad\text{and}\quad a-bd+2(1-a)\rho(x)\leq 0.$$

The solution to this problem is

$$a^*=\frac{-(d+2\alpha)\rho(x)}{\alpha-(d+2\alpha)\rho(x)}\quad\text{and}\quad b^*=\frac{-\rho(x)}{\alpha-(d+2\alpha)\rho(x)}$$

for which

$$a^*-b^*d=\frac{-2\alpha\rho(x)}{\alpha-(d+2\alpha)\rho(x)}.$$

The optimal convergence rate for our estimator in this case is therefore

$$n^{(a^*-b^*d)/2}=n^{-\alpha\rho(x)/(\alpha-(d+2\alpha)\rho(x))}.$$

- If $\gamma_Y(x)<0$ and $\gamma_C(x)<0$, then under the Hölder conditions (4), (5) and (7), hypothesis (15) shall be satisfied if (see (8))

$$\sqrt{kh^d}h^\alpha\log(nh^d)\to 0\quad\text{and}\quad\sqrt{kh^d}\frac{(n/k)^{-\gamma_T(x)}}{L_{U_T}(n/k|x)}h^\beta\to 0\quad\text{as}\quad n\to\infty.$$

  Besides, under assumption (13) and the Hölder conditions (11) and (14), Proposition 1 gives that hypothesis (16) is implied by

$$\sqrt{k_x}\left[|b_Y((U_T(\infty|x)-U_T(n_x/k_x|x))^{-1}|x)|\vee|b_C((U_T(\infty|x)-U_T(n_x/k_x|x))^{-1}|x)|\right]\to 0$$

or, equivalently,

$$\begin{aligned} \sqrt{kh^d}\left(\frac{n}{k}\right)^{\gamma_C(x)\rho_Y(x)/(\gamma_Y(x)+\gamma_C(x))} \mathcal{L}_Y(n/k|x) &\to 0 \\ \text{and } \sqrt{kh^d}\left(\frac{n}{k}\right)^{\gamma_Y(x)\rho_C(x)/(\gamma_Y(x)+\gamma_C(x))} \mathcal{L}_C(n/k|x) &\to 0 \end{aligned}$$

as $n \to \infty$, where $\mathcal{L}_C(\cdot|x)$ and $\mathcal{L}_Y(\cdot|x)$ are slowly varying functions at infinity. Recalling the bias condition $\sqrt{kh^d}\, Q_T(n/k|x) \to 0$ as $n \to \infty$, the problem thus consists in maximizing the quantity $a - bd$ under the constraints $a \in (0, 1)$, $a - bd \geq 0$,

$$\begin{aligned} a - b(d + 2\alpha) &\leq 0, \\ a - 2(1-a)\gamma_T(x) - b(d + 2\beta) &\leq 0, \\ a - bd + 2(1-a)\frac{\gamma_C(x)\rho_Y(x)}{\gamma_Y(x) + \gamma_C(x)} &\leq 0, \\ a - bd + 2(1-a)\frac{\gamma_Y(x)\rho_C(x)}{\gamma_Y(x) + \gamma_C(x)} &\leq 0 \\ \text{and } a - bd + 2(1-a)\rho_T'(x) &\leq 0. \end{aligned}$$

Assume now that the functions γ_Y and γ_C are at least as regular as $U_T(\infty|\cdot)$, namely that $\beta \leq \alpha$. In this case, since $\gamma_T(x) < 0$, the constraints reduce to $a \in (0, 1)$, $a - bd \geq 0$,

$$\begin{aligned} a - bd + 2(1-a)\rho(x) &\leq 0 \\ \text{and } a - 2(1-a)\gamma_T(x) - b(d + 2\beta) &\leq 0 \end{aligned}$$

where

$$\rho(x) := \max\left(\rho_T'(x), \frac{\gamma_C(x)\rho_Y(x)}{\gamma_Y(x) + \gamma_C(x)}, \frac{\gamma_Y(x)\rho_C(x)}{\gamma_Y(x) + \gamma_C(x)}\right) \leq 0.$$

The solution to this problem is

$$a^* = \frac{-(d + 2\beta)\rho(x) - d\gamma_T(x)}{\beta - (d + 2\beta)\rho(x) - d\gamma_T(x)} \quad \text{and} \quad b^* = \frac{-\rho(x) - \gamma_T(x)}{\beta - (d + 2\beta)\rho(x) - d\gamma_T(x)}$$

for which

$$a^* - b^*d = \frac{-2\beta\rho(x)}{\beta - (d + 2\beta)\rho(x) - d\gamma_T(x)}.$$

The optimal convergence rate for our estimator in this case is then

$$n^{(a^* - b^*d)/2} = n^{-\beta\rho(x)/(\beta - (d+2\beta)\rho(x) - d\gamma_T(x))}.$$

5 Simulation study

In this paragraph, we carry out a simulation study to get a grasp of how our estimator behaves in a finite sample situation. We consider the case $E = \mathbb{R}$ equipped with the standard Euclidean norm and a covariate X which is uniformly distributed on $[0,1] \subset E$. Moreover, we let $\gamma_Y : [0,1] \to \mathbb{R}$ and $\gamma_C : [0,1] \to \mathbb{R}$ be the positive functions defined by

$$\forall x \in [0,1],\ \gamma_Y(x) = \frac{2}{3} + \frac{1}{6}\sin(2\pi x) \ \text{ and } \ \gamma_C(x) = 5 + \frac{1}{3}\sin(2\pi x).$$

We shall now give details about the different models we use in this finite-sample study as far as the distribution of (Y, C) given $X = x$ is concerned.

5.1 The models

Fréchet-Fréchet case Our first model is

$$\forall t > 0,\ \overline{F}_Y(t|x) = \left(1 + t^{-\rho/\gamma_Y(x)}\right)^{1/\rho} \ \text{ and } \ \overline{F}_C(t|x) = \left(1 + t^{-\rho/\gamma_C(x)}\right)^{1/\rho}$$

where the parameter ρ is chosen to be independent of x, in the set $\{-1.5, -1, -0.5\}$. In particular, Y and C given $X = x$ are Burr type XII distributed. In this case, $\overline{F}_Y(\cdot|x)$ and $\overline{F}_C(\cdot|x)$ both belong to the Fréchet DA for every $x \in [0,1]$ with respective conditional extreme-value indices $\gamma_Y(x)$ and $\gamma_C(x)$. Finally, the conditional percentage p of censoring in the right tail is such that $0.86 \leq p(x) \leq 0.91$ for all $x \in [0,1]$. We now examine the validity of the second-order condition (M_2). Let us first recall that if a continuous and strictly increasing survival function $\overline{G}$ is such that

$$\overline{G}(t) = t^{-1/\gamma}\left(C_1 + D_1 t^{\rho/\gamma} + \mathrm{o}\left(t^{\rho/\gamma}\right)\right) \ \text{ as } \ t \to \infty$$

where $\gamma > 0$, $\rho < 0$ and C_1, D_1 are nonzero constants, it is straightforward to show that the inverse V of $1/\overline{G}$ is such that

$$V(z) = z^{\gamma}\left(C_2 + D_2 z^{\rho} + \mathrm{o}\left(z^{\rho}\right)\right) \ \text{ as } \ z \to \infty$$

where C_2, D_2 are nonzero constants. The related distribution then satisfies the second-order condition with a second-order parameter equal to ρ. Here, the function $\overline{F}_T(\cdot|x) = \overline{F}_Y(\cdot|x)\overline{F}_C(\cdot|x)$ is continuous and strictly increasing and

$$\overline{F}_T(t|x) = t^{-1/\gamma_T(x)}\left(1 + \frac{1}{\rho}t^{\rho/\gamma_C(x)} + \mathrm{o}\left(t^{\rho/\gamma_C(x)}\right)\right) \ \text{ as } \ t \to \infty$$

because $\gamma_Y(x) < \gamma_C(x)$, so that the second-order condition (M_2) is satisfied in this example, with conditional second-order parameter $\rho_T(x) = \rho\gamma_T(x)/\gamma_C(x) = \rho\gamma_Y(x)/(\gamma_Y(x) + \gamma_C(x))$.

Weibull-Weibull case The second model is

$$\begin{aligned}
\forall t \in [0, g(x)],\ \overline{F}_Y(t|x) &= \frac{\Gamma(2/\gamma_Y(x))}{\Gamma^2(1/\gamma_Y(x))}\int_{t/g(x)}^{1} v^{1/\gamma_Y(x)-1}(1-v)^{1/\gamma_Y(x)-1}dv \\
\text{and } \ \overline{F}_C(t|x) &= \frac{\Gamma(2/\gamma_C(x))}{\Gamma^2(1/\gamma_C(x))}\int_{t/g(x)}^{1} v^{1/\gamma_C(x)-1}(1-v)^{1/\gamma_C(x)-1}dv
\end{aligned}$$

where $\Gamma : (0,\infty) \to \mathbb{R}$ is Euler's Gamma function:

$$\forall z > 0,\ \Gamma(z) = \int_0^{\infty} e^{-t}t^{z-1}dt$$

and the conditional right endpoint function g is defined by

$$\forall x \in [0,1],\ g(x) = 1 - c + 8cx(1-x)$$

with $c \in \{0.1, 0.2, 0.3\}$. Here, given $X = x$, $Y/g(x)$ is a Beta$(1/\gamma_Y(x), 1/\gamma_Y(x))$ random variable and $C/g(x)$ is a Beta$(1/\gamma_C(x), 1/\gamma_C(x))$ random variable. Especially, Y and C given $X = x$ belong to the Weibull DA, with common conditional right endpoint $g(x)$, respective conditional extreme-value indices $-\gamma_Y(x)$ and $-\gamma_C(x)$ and the conditional percentage p of censoring in the right tail being such that $0.86 \leq p(x) \leq 0.91$ for all $x \in [0,1]$. To check that the second-order condition holds, we notice again that if a continuous and strictly increasing survival function $\overline{G}$ is such that

$$\overline{G}(t) = (\theta - t)^{-1/\gamma} \left(C_1 + D_1(\theta - t)^{\rho/\gamma} + \mathrm{o}\left((\theta - t)^{\rho/\gamma} \right) \right) \quad \text{as} \quad t \uparrow \theta$$

where γ, $\rho < 0$, C_1 and D_1 are both nonzero and θ is a constant, it is easy to prove that the inverse V of $1/\overline{G}$ is such that

$$\theta - V(z) = z^{\gamma} \left(C_2 + D_2 z^{\rho} + \mathrm{o}\left(z^{\rho} \right) \right) \quad \text{as} \quad z \to \infty$$

where C_2, D_2 are nonzero constants. The related distribution then satisfies the second-order condition with a second-order parameter equal to ρ. Here, the function $\overline{F}_T(\cdot|x) = \overline{F}_Y(\cdot|x)\overline{F}_C(\cdot|x)$ is continuous, strictly increasing and

$$\overline{F}_T(t|x) \propto (g(x) - t)^{-1/\gamma_T(x)} \left(1 + \left[\frac{1 - \gamma_Y^{-1}(x)}{1 + \gamma_Y(x)} + \frac{1 - \gamma_C^{-1}(x)}{1 + \gamma_C(x)} \right] \left[1 - \frac{t}{g(x)} \right] (1 + \mathrm{o}(1)) \right)$$

as $t \uparrow g(x)$, see the asymptotic expansion of the cdf of a Beta random variable in [2], p.68. It is easy to see that

$$\forall x \in [0,1],\ \frac{1 - \gamma_Y^{-1}(x)}{1 + \gamma_Y(x)} + \frac{1 - \gamma_C^{-1}(x)}{1 + \gamma_C(x)} \neq 0$$

except for two values of x which are approximately equal to $x_1 = 0.1661$ and $x_2 = 0.3339$. For x different from x_1 and x_2, the second-order condition (M_2) is thus satisfied, with conditional second-order parameter $\rho_T(x) = \gamma_T(x)$. When $x \in \{x_1, x_2\}$, straightforward computations entail

$$\begin{aligned}
\overline{F}_T(t|x) \quad \propto \quad & (g(x) - t)^{-1/\gamma_T(x)} \left(1 + \left[\frac{(1 - \gamma_Y^{-1}(x))(2 - \gamma_Y^{-1}(x))}{2(1 + 2\gamma_Y(x))} + \frac{(1 - \gamma_C^{-1}(x))(2 - \gamma_C^{-1}(x))}{2(1 + 2\gamma_C(x))} \right. \right. \\
& \left. \left. + \frac{(1 - \gamma_Y^{-1}(x))(1 - \gamma_C^{-1}(x))}{(1 + \gamma_Y(x))(1 + \gamma_C(x))} \right] \left[1 - \frac{t}{g(x)} \right]^2 (1 + \mathrm{o}(1)) \right)
\end{aligned}$$

as $t \uparrow g(x)$; the coefficient before the $[1 - t/g(x)]^2$ term can be shown to be nonzero for $x \in \{x_1, x_2\}$, so that the second-order condition (M_2) is satisfied with conditional second-order parameter $\rho_T(x) = 2\gamma_T(x)$.

Gumbel-Gumbel case The third model is

$$\forall t > 0,\ \overline{F}_Y(t|x) = \overline{F}_C(t|x) = \frac{2}{1 + \exp(q(x)t)}$$

where q is the function defined by

$$\forall x \in [0,1],\ q(x) = 1 + \frac{1}{2}\sin(2\pi x).$$

In this model, $q(x)Y$ and $q(x)C$ given $X = x$ have a common logistic distribution, which is an example of distribution belonging to the Gumbel DA. Here the function p is constant equal to $1/2$. To check that the second-order condition holds here, we check instead the second-order von Mises condition (see [19], p.49): remark that since Y and C have the same conditional distribution, we have $U_T(z|x) = U_Y(\sqrt{z}|x)$. This entails

$$\frac{zU_T''(z|x)}{U_T'(z|x)} + 1 = \frac{1}{2}\left(\frac{\sqrt{z}U_Y''(\sqrt{z}|x)}{U_Y'(\sqrt{z}|x)} + 1\right).$$

Because $U_Y(z|x) = [q(x)]^{-1}\log(2z-1)$ we get

$$\frac{zU_T''(z|x)}{U_T'(z|x)} + 1 = -\frac{1}{2(2\sqrt{z}-1)}$$

and we may thus apply Theorem 2.3.12 p.49 in [19] to obtain that the conditional distribution of T satisfies the second-order condition (M_2) with $A_T(z|x)$ being the right-hand side of the above equation, yielding $\rho(x) = -1/2$ for every x.

To see how the estimator behaves in uncensored cases, namely when inference in the extremes is possible but the correction due to the presence of censoring is not needed, we also consider the three models below where the right tail of Y is much lighter than that of C:

Gumbel-Fréchet case In this model,

$$\forall t > 0,\ \overline{F}_Y(t|x) = \frac{2}{1 + \exp(q(x)t)} \quad \text{and} \quad \overline{F}_C(t|x) = \left(1 + t^{-\rho/\gamma_C(x)}\right)^{1/\rho}$$

where $q(x) = 1 + 0.5\sin(2\pi x)$ and $\rho = -1$. Here, given $X = x$, $q(x)Y$ has a logistic distribution and C is Burr type XII distributed.

Weibull-Fréchet case Here,

$$\begin{aligned}
\forall t > 0,\ \overline{F}_Y(t|x) &= \frac{\Gamma(2/\gamma_Y(x))}{\Gamma^2(1/\gamma_Y(x))}\int_{t/g(x)}^{1} v^{1/\gamma_Y(x)-1}(1-v)^{1/\gamma_Y(x)-1}dv \\
\text{and}\ \overline{F}_C(t|x) &= \left(1 + t^{-\rho/\gamma_C(x)}\right)^{1/\rho}
\end{aligned}$$

where $g(x) = 1-c+8cx(1-x)$, $c = 0.1$ and $\rho = -1$. Here, given $X = x$, $Y/g(x)$ is $\mathrm{Beta}(1/\gamma_Y(x), 1/\gamma_Y(x))$ distributed and C is Burr type XII distributed.

Weibull-Gumbel case In the final model,

$$\forall t>0,\ \overline{F}_Y(t|x) = \frac{\Gamma(2/\gamma_Y(x))}{\Gamma^2(1/\gamma_Y(x))}\int_{t/g(x)}^1 v^{1/\gamma_Y(x)-1}(1-v)^{1/\gamma_Y(x)-1}dv$$
$$\text{and } \overline{F}_C(t|x) = \frac{2}{1+\exp(q(x)t)}$$

where $g(x)=1-c+8cx(1-x)$, $c=0.1$ and $q(x)=1+0.5\sin(2\pi x)$. Here, given $X=x$, $Y/g(x)$ is Beta$(1/\gamma_Y(x),1/\gamma_Y(x))$ distributed and $q(x)C$ has a logistic distribution.

5.2 Selecting the tuning parameters k_x and h

Our goal is to estimate the conditional extreme-value index γ_Y on a grid of points $\{x_1,\ldots,x_M\}$ of $[0,1]$. To this aim, two parameters have to be chosen: the bandwidth h and the number of log-spacings k_x. We adapt a selection procedure that was introduced in [14]:

1) For every bandwidth h in a grid $\{h_1,\ldots,h_P\}$ of possible values of h, we make a preliminary choice of k_x. Let $\widehat{\gamma}_{i,j}(k)=\widehat{\gamma}_{Y,n}(x_i,k,h_j)$ and $\lfloor\cdot\rfloor$ denote the floor function: for each $i\in\{1,\ldots,M\}$, $j\in\{1,\ldots,P\}$ and $k\in\{q_{i,j}+1,\ldots,N_n(x_i,h_j)-q_{i,j}\}$, where $q_{i,j}=\lfloor N_n(x_i,h_j)/10\rfloor\vee 1$, we introduce the set $E_{i,j,k}=\{\widehat{\gamma}_{i,j}(\ell),\ \ell\in\{k-q_{i,j},\ldots,k+q_{i,j}\}\}$. We compute the standard deviation $\Sigma_{i,j}(k)$ of the set $E_{i,j,k}$ for every possible value of k and we record the number $K_{i,j}$ for which this standard deviation reaches its first local minimum and is less than its average value. Namely, $K_{i,j}=q_{i,j}+1$ if $\Sigma_{i,j}$ is increasing, $K_{i,j}=N_n(x_i,h_j)-q_{i,j}$ if $\Sigma_{i,j}$ is decreasing and

$$K_{i,j} = \min\left\{k \text{ such that } \Sigma_{i,j}(k)\le\Sigma_{i,j}(k-1)\wedge\Sigma_{i,j}(k+1)\right.$$
$$\left.\text{and } \Sigma_{i,j}(k)\le\frac{1}{N_n(x_i,h_j)-2q_{i,j}}\sum_{l=q_{i,j}+1}^{N_n(x_i,h_j)-q_{i,j}}\Sigma_{i,j}(l)\right\}$$

otherwise, where we extend $\Sigma_{i,j}$ by setting $\Sigma_{i,j}(q_{i,j})=\Sigma_{i,j}(q_{i,j}+1)$ and $\Sigma_{i,j}(N_n(x_i,h_j)-q_{i,j}+1)=\Sigma_{i,j}(N_n(x_i,h_j)-q_{i,j})$. We then select the value $k_{i,j}$ such that $\widehat{\gamma}_{i,j}(k_{i,j})$ is the median of the set $E_{i,j,K_{i,j}}$.

The main idea of the first part of this procedure is that, for a given point x_i and a given bandwidth h_j, the number of order statistics is chosen in the first reasonable region of stability of the Hill plot related to the function $k\mapsto\widehat{\gamma}_{Y,n}(x_i,k,h_j)$.

2) We now select the bandwidth h: let q' be a positive integer such that $2q'+1<P$. For each $i\in\{1,\ldots,M\}$ and $j\in\{q'+1,\ldots,P-q'\}$, let $F_{i,j}=\{\widehat{\gamma}_{i,\ell}(k_{i,\ell}),\ \ell\in\{j-q',\ldots,j+q'\}\}$ and compute the standard deviation $\sigma_i(j)$ of $F_{i,j}$. Our objective function is then the average of these quantities over the grid $\{x_1,\ldots,x_M\}$:

$$\overline{\sigma}(j)=\frac{1}{M}\sum_{i=1}^M\sigma_i(j).$$

We next record the integer j^* such that $\overline{\sigma}(j^*)$ is the first local minimum of the application $j \mapsto \overline{\sigma}(j)$ which is less than the average value of $\overline{\sigma}$. In other words, $j^* = q'+1$ if $\overline{\sigma}$ is increasing, $j^* = P - q'$ if $\overline{\sigma}$ is decreasing and

$$j^* = \min\left\{ j \text{ such that } \overline{\sigma}(j) \leq \overline{\sigma}(j-1) \wedge \overline{\sigma}(j+1) \text{ and } \overline{\sigma}(j) \leq \frac{1}{P-2q'} \sum_{l=q'+1}^{P-q'} \overline{\sigma}(l) \right\}$$

otherwise, where we extend $\overline{\sigma}$ by setting $\overline{\sigma}(q') = \overline{\sigma}(q'+1)$ and $\overline{\sigma}(P-q'+1) = \overline{\sigma}(P-q')$. The selected bandwidth is then independent of x and is given by $h^* = h_{j^*}$.

In doing so, we require that h^* be not too large, to ensure that the computation of our estimator is carried out only using covariates which are close to x, and the estimation carried out for bandwidths in a neighborhood of h^* is reasonably stable. The selected number of log-spacings is thus given, for $x = x_i$, by $k^*_{x_i} = k_{i,j^*}$.

We choose here to estimate the conditional extreme-value index on a grid of $M = 50$ evenly spaced points in $[0, 1]$. Regarding the selection procedure, we test $P = 25$ evenly spaced values of h ranging from 0.05 to 0.25 and we set $q' = 1$.

5.3 Results

We give in Table 1 the empirical mean squared errors (MSEs) of our estimator, averaged over the M points of the grid, for $N = 100$ independent samples of size $n = 1000$, along with the minimal and maximal MSEs obtained. One can see that in the Fréchet-Fréchet case, the MSE of our estimator increases as $|\rho|$ approaches 0: this is not surprising since the conditional second-order parameter of T, known to play a major role in the performance of the estimators of the extreme-value index, is proportional to ρ in this case. The Weibull-Weibull case seems to show that the quality of the estimates does not depend on the value of the common endpoint, and this could be expected as well since we know that how the estimator performs should only depend on the value of the second-order parameter and of the censoring percentage in the extremes. Finally, in the cases tested here, the estimator performs much better in the uncensored cases than in the censored cases at the finite-sample level. Some illustrations are given in Figures 1 and 2, where the estimates corresponding to the 5% quantile, median and 95% quantile of the MSE are represented in each case for our estimator.

6 Real data example

In this section, we introduce a medical data set, provided by Dr P. J. Solomon and the Australian National Centre in HIV Epidemiology and Clinical Research; see Ripley and Solomon [24], Venables and Ripley [27] and the data set `aids2`, part of the package MASS in R. In the context of extreme value analysis, this data set was considered by [11] and [23]. The data set contains information collected after a follow-up study on 2843 patients diagnosed with AIDS before July 1st, 1991. Especially, for each patient, the data set gives his/her age at the time of diagnosis and, if the patient died before the end of the study, his/her date of death. There are only 89 female patients in this study, so we chose to retain the 2754 male patients of the data set. Our variable of interest is the survival time Y of a patient which is randomly right-censored as is usually the case in such follow-up studies. The covariate we consider is the age of a patient at the time of diagnosis. A scatterplot of the data is given in Figure 3.

Our first goal is to provide an estimate of the conditional extreme-value index of Y using our estimator. A look at the scatterplot shows that data for patients aged either less than 20 or more than 65 when diagnosed with AIDS is very scarce, so we focus on patients aged between $x_{\min} = 20$ and $x_{\max} = 65$. We use the selection procedure detailed in Section 5.2: the bandwidth h is chosen among $h_1 \leq \cdots \leq h_{25}$ where the h_i are evenly spaced and

$$h_1 = 0.05(x_{\max} - x_{\min}) \text{ and } h_{25} = 0.25(x_{\max} - x_{\min}).$$

This leads us to choose $h^* = 3.75$. The estimate of the conditional extreme-value index γ_Y on 25 evenly spaced points in $[x_{\min}, x_{\max}]$ is represented on Figure 4.

This estimate is only a first step in the assessment of the tail heaviness of the conditional distribution of Y given $X = x$, however. A further interesting step is to estimate conditional extreme quantiles of this distribution, where the conditional quantile function $q_Y(\cdot|x)$ is defined in terms of the generalized inverse of $\overline{F}_Y(\cdot|x)$:

$$q_Y(\varepsilon|x) = \inf\{t \in \mathbb{R} \,|\, \overline{F}_Y(t|x) \leq \varepsilon\}.$$

Note at this point that $q_Y(\varepsilon|x)$ is the $(1-\varepsilon)-$conditional quantile of Y in the usual sense. We propose an adaptation of the extreme quantile estimator of [11], which is itself an adaptation of the classical extreme quantile estimator, see for instance Theorem 4.3.1 in [19], p.134. We let $\widehat{\overline{F}}_{Y,n}(\cdot, h|x)$ be the straightforward conditional adaptation of the Kaplan-Meier estimator for the csf of Y given $X = x$ (see Beran [5]). Besides, given $N_n(x,h) = l$, we set for $k_x \in \{1, \ldots, l-1\}$

$$\widehat{a}_n(x, k_x, h) = \mathcal{T}_{l-k_x,l} \frac{\widehat{\gamma}_{T,n,+}(x, k_x, h)(1 - \widehat{\gamma}_{T,n,-}(x, k_x, h))}{\widehat{p}_n(x, k_x, h)}$$

and 0 otherwise. An estimator of the conditional extreme quantile $q_Y(\varepsilon|x)$, where ε is a small positive number, is then

$$\widehat{q}_{Y,n}(\varepsilon, x, k_x, h) = \mathcal{T}_{l-k_x,l} + \widehat{a}_n(x, k_x, h) D_{\widehat{\gamma}_{T,n}(x,k_x,h)} \left(\widehat{\overline{F}}_{Y,n}(\mathcal{T}_{l-k_x,l}, h|x)/\varepsilon \right)$$

if $k_x \in \{1, \ldots, l-1\}$ and 0 otherwise, where the function D was introduced in (1). In our case, we set $h = h^*$; for $x \in [x_{\min}, x_{\max}]$, the number of log-spacings k_x is chosen by applying the first step of the selection procedure introduced in Section 5.2.

We give some results on Figure 5, where estimates of the extreme quantile curve $x \mapsto \widehat{q}_{Y,n}(\varepsilon, x, k_x^*, h^*)$ are represented for an exceedance level $\varepsilon \in \{0.01, 0.005, 0.002, 0.001\}$. One can see on this figure that these estimates are fairly stable for patients aged between 20 and 53 years and decrease sharply afterwards. This may be interpreted as a consequence of immunosenescence, namely the deterioration of the immune system as age increases. This phenomenon is of course especially critical in the case of AIDS, since HIV targets cells of the immune system; the significant effect of increasing age on survival rates for AIDS has been shown numerous times in the medical literature, see *e.g.* Ripley and Solomon [24], Luo *et al.* [22], Darby *et al.* [9] and Balslev *et al.* [1], among others. Besides, one can see that the estimate of the extreme quantile curve for $\varepsilon = 0.001$ yields, in the range $[20, 53]$, survival times around 13 years and as high as 16 years. This is in line with Figure 1(b) of [11], which does not consider any covariate information and gives a value of this extreme survival time between 15 and 19 years while using a different estimator of the extreme-value index.

7 Proofs

Before giving a proof of Theorem 1, we need some preliminary results. Lemma 1, which is essentially contained in [11], gives a useful representation of $p(x)$.

Lemma 1. *Let Y, C be two independent positive random variables having respective survival functions $\overline{F}_Y$, $\overline{F}_C$, respective pdfs f_Y, f_C and common right endpoint $U(\infty) = U_Y(\infty) = U_C(\infty)$. Define for $t > 0$*

$$p(t) = \frac{d}{dt}\mathbb{P}(Y \le C, Y \wedge C \le t) \Big/ \frac{d}{dt}\mathbb{P}(Y \wedge C \le t)$$

whenever the denominator is nonzero, and $p := \gamma_C/(\gamma_Y + \gamma_C)$ otherwise. Then one has

$$p(t) = \frac{\overline{F}_C(t) f_Y(t)}{\overline{F}_C(t) f_Y(t) + \overline{F}_Y(t) f_C(t)}$$

whenever the denominator is nonzero. In particular, $p(t) \le 1$ for every $t > 0$. If moreover Y and C belong respectively to $\mathcal{D}(G_{\gamma_Y})$ and $\mathcal{D}(G_{\gamma_C})$ and either

- *$\gamma_Y > 0$ and $\gamma_C > 0$;*
- *$\gamma_Y < 0$, $\gamma_C < 0$ and $0 < U(\infty) < \infty$,*

then $p(t) \to p$ as $t \to U(\infty)$.

Lemma 2 is a partial generalization of Lemma 1 to the random covariate case.

Lemma 2. *Assume that the functions $(x, t) \mapsto f_Y(t|x)$ and $(x, t) \mapsto f_C(t|x)$ are continuous on $E \times (0, \infty)$. Then given $X \in B(x, h)$, T has pdf*

$$f_{T,h}(t|x) := \mathbb{E}(\overline{F}_C(t|X) f_Y(t|X) + \overline{F}_Y(t|X) f_C(t|X) \,|\, X \in B(x, h))$$

and we have

$$\forall t > 0,\ p_h(t|x) = \frac{\mathbb{E}(\overline{F}_C(t|X) f_Y(t|X) \,|\, X \in B(x, h))}{\mathbb{E}(\overline{F}_C(t|X) f_Y(t|X) \,|\, X \in B(x, h)) + \mathbb{E}(\overline{F}_Y(t|X) f_C(t|X) \,|\, X \in B(x, h))}$$

whenever the denominator is nonzero. In particular, $p_h(t|x) \le 1$ for every $t > 0$.

Proof of Lemma 2. Remark that

$$F_{T,h}(t|x) = \mathbb{P}(Y \le C, Y \le t \,|\, X \in B(x, h)) + \mathbb{P}(C \le Y, C \le t \,|\, X \in B(x, h)).$$

The independence of Y and C given X and Tonelli's theorem yield

$$\begin{aligned}
\mathbb{P}(Y \leq C, Y \leq t \,|\, X \in B(x,h)) &= \mathbb{E}\left(\int_0^t \overline{F}_C(z|X) f_Y(z|X) dz \,|\, X \in B(x,h)\right) \\
&= \int_0^t \mathbb{E}\left(\overline{F}_C(z|X) f_Y(z|X) \,|\, X \in B(x,h)\right) dz \qquad (17) \\
\text{and } \mathbb{P}(C \leq Y, C \leq t \,|\, X \in B(x,h)) &= \mathbb{E}\left(\int_0^t \overline{F}_Y(z|X) f_C(z|X) dz \,|\, X \in B(x,h)\right) \\
&= \int_0^t \mathbb{E}\left(\overline{F}_Y(z|X) f_C(z|X) \,|\, X \in B(x,h)\right) dz.
\end{aligned}$$

The regularity hypotheses on f_Y and f_C make it clear that both of the above integrands are continuous as functions of z, so that $F_{T,h}(\cdot|x)$ has a continuous derivative which is

$$\frac{d}{dt} F_{T,h}(t|x) = \mathbb{E}(\overline{F}_C(t|X) f_Y(t|X) + \overline{F}_Y(t|X) f_C(t|X) \,|\, X \in B(x,h)) = f_{T,h}(t|x). \qquad (18)$$

This is the first desired result. Moreover,

$$\mathbb{P}(\delta = 1, T \leq t \,|\, X \in B(x,h)) = \mathbb{P}(Y \leq C, Y \leq t \,|\, X \in B(x,h)).$$

From (17), we get

$$\frac{d}{dt}\mathbb{P}(\delta = 1, T \leq t \,|\, X \in B(x,h)) = \mathbb{E}\left(\overline{F}_C(t|X) f_Y(t|X) \,|\, X \in B(x,h)\right). \qquad (19)$$

Combining (18) and (19) concludes the proof. ■

We then state a couple of useful technical results. The first one gives the conditional distribution of the random pairs $(\mathcal{T}_i, \Delta_i)$.

Lemma 3. *Given $N_n(x,h) = l \geq 1$, the random pairs $(\mathcal{T}_i, \Delta_i)$, $1 \leq i \leq l$, are independent and identically distributed random variables whose common distribution is that of (T, δ) given $X \in B(x,h)$.*

Proof of Lemma 3. The proof of this result is similar to that of Lemma 2 in [26]: if $(t_1, \ldots, t_l) \in \mathbb{R}^l$ and $(d_1, \ldots, d_l) \in \{0,1\}^l$, then since the random vectors (X_i, T_i, δ_i) have the same distribution, it holds that

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{l} \{\mathcal{T}_i \leq t_i, \Delta_i = d_i\}, N_n(x,h) = l\right) &= \binom{n}{l} \mathbb{P}\left(\bigcap_{i=1}^{l} \{T_i \leq t_i, \delta_i = d_i, X_i \in B(x,h)\}\right) \\
&\times \prod_{i=l+1}^{n} \mathbb{P}\left(X_i \notin B(x,h)\right).
\end{aligned}$$

The independence of the random pairs (X_i, T_i, δ_i), $i = 1, \ldots, n$ entails that the above probability is

$$\prod_{i=1}^{l} \mathbb{P}(T \leq t_i, \delta = d_i \,|\, X \in B(x,h)) \times \left[\binom{n}{l} \prod_{i=1}^{l} \mathbb{P}\left(X_i \in B(x,h)\right) \prod_{i=l+1}^{n} \mathbb{P}\left(X_i \notin B(x,h)\right)\right].$$

Since $N_n(x,h)$ is a binomial random variable with parameters n and $\mathbb{P}(X \in B(x,h))$, the result follows. ■

The next lemma, whose proof can be found in [26], is a pivotal technical tool for the proofs of Theorems 1 and 2.

Lemma 4. *Let (S_n) be a sequence of random variables. Assume that there exist a triangular array of events $(A_{ij})_{0\le j\le i}$ and a sequence of non-empty sets (I_n) contained in $\{1,\dots,n\}$ such that*

- *for every n the A_{nl}, $0\le l\le n$, have positive probability, are pairwise disjoint and*
$$\sum_{l=0}^{n}\mathbb{P}(A_{nl})=1;$$
- *it holds that $\sum_{l\in I_n}\mathbb{P}(A_{nl})\to 1$ as $n\to\infty$.*

If one has for every $\varepsilon>0$
$$\sup_{l\in I_n}\mathbb{P}(|S_n|>\varepsilon|A_{nl})\to 0 \quad as \quad n\to\infty,$$
then $S_n \xrightarrow{\mathbb{P}} 0$ as $n\to\infty$.

This result will be applied in the following way: remark that since $N_n(x,h)$ is a binomial random variable with parameters n and $\mathbb{P}(X\in B(x,h))$, it is a consequence of Chebyshev's inequality that for all $\eta\in(0,1)$,
$$\sqrt{n_x^{1-\eta}}\left|\frac{N_n(x,h)}{n_x}-1\right| \xrightarrow{\mathbb{P}} 0 \;\text{ as }\; n\to\infty.$$
If $I_x := \mathbb{N}\cap\left[\left(1-n_x^{-1/4}\right)n_x,\left(1+n_x^{-1/4}\right)n_x\right]$ – this notation will be used in the remainder of Section 7 – then this entails
$$\sum_{l\in I_x}\mathbb{P}(N_n(x,h)=l)\to 1 \;\text{ as }\; n\to\infty.$$

The final lemma, contained in [26], makes it possible to understand a bit more about the asymptotic behavior of certain random variables which appear in our proofs.

Lemma 5. *Let W_i, $i\ge 1$ be independent standard Pareto random variables, i.e. having cdf $w\mapsto 1-1/w$ on $(1,\infty)$. Assume that $n_x\to\infty$, $k_x\to\infty$ and $k_x/n_x\to 0$ as $n\to\infty$. Then for every $\varepsilon>0$ it holds that*
$$\sup_{l\in I_x}\mathbb{P}\left(\left|\frac{k_x}{l}W_{l-k_x,l}-1\right|>\varepsilon\right)\to 0 \quad as \quad n\to\infty.$$

We may now prove Theorem 1.

Proof of Theorem 1. Write
$$\widehat{\gamma}_{Y,n}(x,k_x,h)-\gamma_Y(x)=\frac{1}{\widehat{p}_n(x,k_x,h)}\left[(\widehat{\gamma}_{T,n}(x,k_x,h)-\gamma_T(x))-\frac{\gamma_T(x)}{p(x)}(\widehat{p}_n(x,k_x,h)-p(x))\right].$$

Following [11], we note that if V is a standard uniform random variable which is independent of (X,Y,C), then:

$$\begin{aligned}\mathbb{P}(V \le p_h(T|x), T \le t_0 \mid X \in B(x,h)) &= \int_0^{t_0} p_h(t|x) f_{T,h}(t|x)dt \\ &= \mathbb{P}(\delta = 1, T \le t_0 \mid X \in B(x,h))\end{aligned}$$

so that given $X \in B(x,h)$, the random pairs $(T, \mathbb{I}_{\{V \le p_h(T|x)\}})$ and (T,δ) have the same distribution. Consequently, if V_i, $i \ge 1$ is an independent sequence of standard uniform random variables which are independent of the (X_i, Y_i, C_i), then given $N_n(x,h) = l$, it is a consequence of Lemma 3 that the distribution of $(\widehat{\gamma}_{T,n}(x,k_x,h), \widehat{p}_n(x,k_x,h))$ is that of $(\widehat{\gamma}_{T,n}(x,k_x,h), \widetilde{p}_n(x,k_x,h))$, with

$$\widetilde{p}_n(x,k_x,h) := \frac{1}{k_x} \sum_{i=1}^{k_x} \mathbb{I}_{\{V_{[l-i+1:l]} \le p_h(\mathcal{T}_{l-i+1,l}|x)\}}$$

if $k_x \in \{1,\ldots,l-1\}$ and 0 otherwise, where $V_{[1:l]},\ldots,V_{[l:l]}$ are the order statistics induced by $\mathcal{T}_{1,l},\ldots,\mathcal{T}_{l,l}$. Moreover, since the V_i, $i \ge 1$ are standard uniform variables independent of the (X_i,Y_i,C_i), so are the $V_{[i:l]}$, $1 \le i \le l$. Introducing, given $N_n(x,h) = l$, the quantity

$$\overline{p}_n(x,k_x,h) := \frac{1}{k_x} \sum_{i=1}^{k_x} \mathbb{I}_{\{V_i \le p_h(\mathcal{T}_{l-i+1,l}|x)\}}$$

if $k_x \in \{1,\ldots,l-1\}$ and 0 otherwise, we obtain

$$\widehat{\gamma}_{Y,n}(x,k_x,h) - \gamma_Y(x) \stackrel{d}{=} \frac{1}{\overline{p}_n(x,k_x,h)} \left[(\widehat{\gamma}_{T,n}(x,k_x,h) - \gamma_T(x)) - \frac{\gamma_T(x)}{p(x)} (\overline{p}_n(x,k_x,h) - p(x)) \right].$$

It is thus enough to show the consistency of $\widehat{\gamma}_{T,n}(x,k_x,h)$ and $\overline{p}_n(x,k_x,h)$. The consistency of the former quantity is an immediate consequence of Theorem 1 in [26]. To prove the consistency of $\overline{p}_n(x,k_x,h)$, note that

$$\overline{p}_n(x,k_x,h) - p(x) = \left[\frac{B_{k_x}}{k_x} - p(x) \right] - S_{n,1} + S_{n,2}$$

where

$$B_{k_x} = \sum_{i=1}^{k_x} \mathbb{I}_{\{V_i \le p(x)\}}, \tag{20}$$

$$S_{n,1} = \mathbb{I}_{\{N_n(x,h) \le k_x\}} \sum_{i=1}^{k_x} \mathbb{I}_{\{V_i \le p(x)\}} \tag{21}$$

$$\text{and } S_{n,2} = \sum_{l=k_x+1}^{n} \left[\frac{1}{k_x} \sum_{i=1}^{k_x} \mathbb{I}_{\{V_i \le p_h(\mathcal{T}_{l-i+1,l}|x)\}} - \mathbb{I}_{\{V_i \le p(x)\}} \right] \mathbb{I}_{\{N_n(x,h)=l\}}. \tag{22}$$

As a consequence, B_{k_x} is a binomial random variable with parameters k_x and $p(x)$ which is independent of $\widehat{\gamma}_{T,n}(x,k_x,h)$ and Tchebychev's inequality entails

$$\overline{p}_n(x,k_x,h) - p(x) = -S_{n,1} + S_{n,2} + o_{\mathbb{P}}(1) \quad \text{as} \quad n \to \infty.$$

Further, for every $\varepsilon > 0$,

$$\mathbb{P}(|S_{n,1}| > \varepsilon) \leq \mathbb{P}(N_n(x,h) \leq k_x) \to 0 \text{ as } n \to \infty$$

so that $S_{n,1} \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. Besides, if W_i, $i \geq 1$ are independent standard Pareto random variables, then the distribution of the random vector $(\mathcal{T}_1, \ldots, \mathcal{T}_l)$ given $N_n(x,h) = l \geq 1$ is the distribution of the random vector $(U_{T,h}(W_1|x), \ldots, U_{T,h}(W_l|x))$, see Lemma 3. Let n be so large that $k_x < \inf I_x$. The equality

$$\forall a, b \in [0,1], \ \mathbb{E}\left|\mathbb{I}_{\{V \leq a\}} - \mathbb{I}_{\{V \leq b\}}\right| = |a - b|$$

valid for every standard uniform random variable V, entails for every $l \in I_x$

$$\begin{aligned}
\mathbb{E}(|S_{n,2}| \,|\, N_n(x,h) = l) &\leq \frac{1}{k_x} \sum_{i=1}^{k_x} \mathbb{E}(|p_h(\mathcal{T}_{l-i+1,l}|x) - p(x)| \,|N_n(x,h) = l) \\
&= \frac{1}{k_x} \sum_{i=1}^{k_x} \mathbb{E}|p_h(U_{T,h}(W_{l-i+1,l}|x)|x) - p(x)|.
\end{aligned}$$

Clearly, for every $\kappa > 0$, if n is so large that

$$\omega\left(p \circ U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) \leq \frac{\kappa}{2}$$

we have by Lemma 2 that

$$\mathbb{E}(|S_{n,2}| \,|\, N_n(x,h) = l) \leq \frac{\kappa}{2} + 2 \sup_{l \in I_x} \mathbb{P}\left(\{W_{l-k_x+1,l} < n_x/(1+\eta)k_x\} \cup \{W_{l,l} > n_x^{1+\eta}\}\right). \tag{23}$$

Lemma 5 entails

$$\sup_{l \in I_x} \mathbb{P}(W_{l-k_x+1,l} < n_x/(1+\eta)k_x) = \sup_{l \in I_x} \mathbb{P}\left(\frac{k_x}{n_x} W_{l-k_x+1,l} - 1 < -\frac{\eta}{1+\eta}\right) \to 0 \text{ as } n \to \infty,$$

and since the W_i are independent standard Pareto random variables, we get

$$\sup_{l \in I_x} \mathbb{P}\left(W_{l,l} > n_x^{1+\eta}\right) = \sup_{l \in I_x} \left[1 - \left(1 - n_x^{-1-\eta}\right)^l\right] \leq 1 - \left(1 - n_x^{-1-\eta}\right)^{3n_x/2} \to 0 \text{ as } n \to \infty.$$

In other words

$$\sup_{l \in I_x} \mathbb{P}(\{W_{l-k_x+1,l} < n_x/(1+\eta)k_x\} \cup \{W_{l,l} > n_x^{1+\eta}\}) \leq \frac{\kappa}{4} \tag{24}$$

for n large enough, so that combining (23) and (24), we find that $\mathbb{E}(|S_{n,2}| \,|\, N_n(x,h) = l) \to 0$ uniformly in $l \in I_x$ as $n \to \infty$. According to Markov's inequality, we have for every $\varepsilon > 0$

$$\sup_{l \in I_x} \mathbb{P}(|S_{n,2}| > \varepsilon \,|\, N_n(x,h) = l) \leq \sup_{l \in I_x} \frac{\mathbb{E}(|S_{n,2}| \,|\, N_n(x,h) = l)}{\varepsilon} \to 0 \text{ as } n \to \infty.$$

Lemma 4 then entails $S_{n,2} \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ and the proof is complete. ■

We proceed by proving the pointwise asymptotic normality of the estimator.

Proof of Theorem 2. Recall from the proof of Theorem 1 the equality

$$\widehat{\gamma}_{Y,n}(x,k_x,h)-\gamma_Y(x)\stackrel{d}{=}\frac{1}{\overline{p}_n(x,k_x,h)}\left[(\widehat{\gamma}_{T,n}(x,k_x,h)-\gamma_T(x))-\frac{\gamma_T(x)}{p(x)}(\overline{p}_n(x,k_x,h)-p(x))\right].$$

The asymptotic normality of $\widehat{\gamma}_{T,n}(x,k_x,h)$,

$$\sqrt{k_x}\,[\widehat{\gamma}_{T,n}(x,k_x,h)-\gamma_T(x)]\xrightarrow{d}\mathcal{N}(0,V(\gamma_T(x))) \tag{25}$$

is contained in Theorem 2 of [26]. We now recall the representation

$$\overline{p}_n(x,k_x,h)-p(x)=\left[\frac{B_{k_x}}{k_x}-p(x)\right]-S_{n,1}+S_{n,2}$$

with B_{k_x}, $S_{n,1}$ and $S_{n,2}$ as in (20), (21) and (22). Note that, from (21), one has for every $\varepsilon>0$

$$\mathbb{P}(\sqrt{k_x}|S_{n,1}|>\varepsilon)\leq\mathbb{P}(N_n(x,h)\leq k_x)\to 0\ \text{ as }\ n\to\infty$$

so that $\sqrt{k_x}|S_{n,1}|\xrightarrow{\mathbb{P}}0$ as $n\to\infty$. Let n be so large that $k_x<\inf I_x$. Let further W_i, $i\geq 1$ be independent standard Pareto random variables which are independent of the V_i and note that, from Lemma 3 and (22), one has given $N_n(x,h)=l\in I_x$:

$$S_{n,2}\stackrel{d}{=}\frac{1}{k_x}\sum_{i=1}^{k_x}\mathbb{I}_{\{V_i\leq p_h(U_{T,h}(W_{l-i+1,l}|x)|x)\}}-\mathbb{I}_{\{V_i\leq p(x)\}}=:S_n'.$$

Further,

$$\begin{aligned}\sqrt{k_x}|S_n'| &\leq 2\sqrt{k_x}\mathbb{I}_{\{W_{l-k_x+1,l}<n_x/(1+\eta)k_x\}\cup\{W_{l,l}>n_x^{1+\eta}\}}\\ &+\sqrt{k_x}\left[\frac{1}{k_x}\sum_{i=1}^{k_x}\left|\mathbb{I}_{\{V_i\leq p_h(U_{T,h}(W_{l-i+1,l}|x)|x)\}}-\mathbb{I}_{\{V_i\leq p(x)\}}\right|\right]\\ &\times\ \mathbb{I}_{\{n_x/(1+\eta)k_x\leq W_{l-k_x+1,l}\leq W_{l,l}\leq n_x^{1+\eta}\}}.\end{aligned}$$

Since the expectation of the second term on the right-hand side of this inequality is

$$\frac{1}{\sqrt{k_x}}\sum_{i=1}^{k_x}\mathbb{E}\left[|p_h(U_{T,h}(W_{l-i+1,l}|x)|x)-p(x)|\mathbb{I}_{\{n_x/(1+\eta)k_x\leq W_{l-k_x+1,l}\leq W_{l,l}\leq n_x^{1+\eta}\}}\right]$$

we may, for every $\kappa>0$, bound it from above by

$$\sqrt{k_x}\omega\left(p\circ U_T,\frac{n_x}{(1+\eta)k_x},n_x^{1+\eta},x,h\right)\leq\frac{\kappa}{2}$$

for n sufficiently large. From (24) and Markov's inequality, we get for every $\varepsilon>0$

$$\sup_{l\in I_x}\mathbb{P}(\sqrt{k_x}|S_{n,2}|>\varepsilon\,|\,N_n(x,h)=l)\leq\kappa$$

if n is large enough. By Lemma 4, this entails $\sqrt{k_x}|S_{n,2}| \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. Consequently

$$\sqrt{k_x}\left[\overline{p}_n(x, k_x, h) - p(x)\right] = \sqrt{k_x}\left[\frac{B_{k_x}}{k_x} - p(x)\right] + o_{\mathbb{P}}(1).$$

Recall from the proof of Theorem 1 that B_{k_x} is a binomial random variable with parameters k_x and $p(x)$ which is independent of $\widehat{\gamma}_{T,n}(x, k_x, h)$. Since

$$\sqrt{k_x}\left[\frac{B_{k_x}}{k_x} - p(x)\right] \xrightarrow{d} \mathcal{N}(0, p(x)(1 - p(x))), \tag{26}$$

as $n \to \infty$, the convergences (25), (26) and Slutsky's lemma entail

$$\sqrt{k_x}\left[\widehat{\gamma}_{Y,n}(x, k_x, h) - \gamma_Y(x)\right] \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{p^2(x)}\left[V(\gamma_T(x)) + \frac{\gamma_T^2(x)}{p(x)}(1 - p(x))\right]\right)$$

as $n \to \infty$, which is the result. ■

The last lemma is the converse statement of Lemma 9 in [26]. It is necessary to prove Proposition 1.

Lemma 6. *Let $\overline{F}$ be a csf on $\mathbb{R}$ and U be the left-continuous inverse of $1/\overline{F}$.*

1. *If $\overline{F}$ is such that*
$$\forall y \in \mathbb{R},\ \overline{F}(y) \in (0, 1) \Rightarrow \forall \delta > 0,\ \overline{F}(y + \delta) < \overline{F}(y)$$
then U is a continuous function on $(1, \infty)$.
2. *If $\overline{F}$ is continuous on $\mathbb{R}$ then U is an increasing function on $(1, \infty)$.*

Proof of Lemma 6. To prove the first statement, pick $\alpha_0 \in (1, \infty)$ and assume that U is not continuous at α_0. In particular, since U is left-continuous and nondecreasing,

$$\lim_{\substack{\alpha \to \alpha_0 \\ \alpha > \alpha_0}} U(\alpha) - U(\alpha_0) > 0.$$

Then necessarily $0 < \overline{F}(U(\alpha_0)) \leq 1/\alpha_0 < 1$. Moreover, the above inequality entails, since U is nondecreasing,

$$\exists \delta > 0,\ \forall \alpha > \alpha_0,\ U(\alpha) > U(\alpha_0) + \delta.$$

Using the definition of the function U, we obtain

$$\forall \alpha > \alpha_0,\ \alpha_0 \leq \frac{1}{\overline{F}(U(\alpha_0))} \leq \frac{1}{\overline{F}(U(\alpha_0) + \delta)} < \alpha.$$

Taking the limit $\alpha \downarrow \alpha_0$ gives $\overline{F}(U(\alpha_0) + \delta) = \overline{F}(U(\alpha_0))$, which is a contradiction.

To show the second statement, assume that α, β are such that $1 < \alpha < \beta$ and $U(\alpha) = U(\beta)$. Then since $\overline{F}$ is right-continuous and nonincreasing, we get

$$\overline{F}(U(\alpha)) = \overline{F}(U(\beta)) \leq \frac{1}{\beta} < \frac{1}{\alpha} \leq \lim_{\substack{t \to U(\alpha) \\ t < U(\alpha)}} \overline{F}(t).$$

Hence $\overline{F}$ is not continuous at $U(\alpha)$, which is a contradiction. ■

Proof of Proposition 1. We start by considering case 1. For n large enough and for every $x' \in B(x,h)$, one has

$$\begin{aligned} \overline{F}_C(t|x')f_Y(t|x') - \frac{1}{\gamma_Y(x)}G(t|x') &= r_Y(t,x,x')G(t|x') \\ \text{and} \ \ \overline{F}_Y(t|x')f_C(t|x') - \frac{1}{\gamma_C(x)}G(t|x') &= r_C(t,x,x')G(t|x') \end{aligned}$$

with $G(t|x') = t^{-1/\gamma_Y(x')-1/\gamma_C(x')-1}L_{\overline{F}_Y}(t|x')L_{\overline{F}_C}(t|x')$,

$$\begin{aligned} r_Y(t,x,x') &= \frac{1}{\gamma_Y(x')} - \frac{1}{\gamma_Y(x)} - b_Y(t|x') \\ \text{and} \ \ r_C(t,x,x') &= \frac{1}{\gamma_C(x')} - \frac{1}{\gamma_C(x)} - b_C(t|x'). \end{aligned}$$

From Lemma 2, we obtain the equality

$$p_h(t|x) = \frac{\dfrac{1}{\gamma_Y(x)} + \dfrac{\mathbb{E}(r_Y(t,x,X)G(t|X) \,|\, X \in B(x,h))}{\mathbb{E}(G(t|X) \,|\, X \in B(x,h))}}{\dfrac{1}{\gamma_Y(x)} + \dfrac{1}{\gamma_C(x)} + \dfrac{\mathbb{E}([r_Y(t,x,X)+r_C(t,x,X)]G(t|X) \,|\, X \in B(x,h))}{\mathbb{E}(G(t|X) \,|\, X \in B(x,h))}}.$$

If we can prove that for $\eta > 0$ small enough

$$\sup_{t \in U_{T,h}(K_{x,\eta}|x)} \sup_{x' \in B(x,h)} (|b_Y| \vee |b_C|)(t|x') = \mathrm{O}(h^\alpha \log n_x \vee \delta_n) \to 0 \tag{27}$$

as $n \to \infty$, with $U_{T,h}(K_{x,\eta}|x)$ being the image of the interval $K_{x,\eta}$ by the function $U_{T,h}(\cdot|x)$, then the fact that $G(\cdot|X)$ is nonnegative shall entail

$$\begin{aligned} \sup_{t \in U_{T,h}(K_{x,\eta}|x)} \left| \frac{\mathbb{E}(r_Y(t,x,X)G(t|X) \,|\, X \in B(x,h))}{\mathbb{E}(G(t|X) \,|\, X \in B(x,h))} \right| &\leq \sup_{t \in U_{T,h}(K_{x,\eta}|x)} \sup_{x' \in B(x,h)} |r_Y(t,x,x')| \\ &= \mathrm{O}(h^\alpha \log n_x \vee \delta_n) \\ \text{and} \ \ \sup_{t \in U_{T,h}(K_{x,\eta}|x)} \left| \frac{\mathbb{E}(r_C(t,x,X)G(t|X) \,|\, X \in B(x,h))}{\mathbb{E}(G(t|X) \,|\, X \in B(x,h))} \right| &\leq \sup_{t \in U_{T,h}(K_{x,\eta}|x)} \sup_{x' \in B(x,h)} |r_C(t,x,x')| \\ &= \mathrm{O}(h^\alpha \log n_x \vee \delta_n) \end{aligned}$$

of which it is a direct consequence that

$$\omega\left(p \circ U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) = \mathrm{O}(h^\alpha \log n_x \vee \delta_n)$$

which is the result. To this end, we start by noting that because (see Lemma 1.2.9 in [19], p.22)

$$\frac{U_T(n_x/k_x|x)}{a_T(n_x/k_x|x)} \to \frac{1}{\gamma_T(x)} \ \text{ as } \ n \to \infty,$$

it is a consequence of (6) and of the mean value theorem that

$$\sup_{z\in K_{x,\eta}} \left| \frac{U_{T,h}(z|x)}{U_T(z|x)} - 1 \right| \to 0 \quad \text{as} \quad n \to \infty.$$

Using the fact that $U_T(\cdot|x)$ is regularly varying at infinity with index $\gamma_T(x) > 0$, we get for n large enough

$$U_{T,h}(K_{x,\eta}|x) \subset U_T(K_{x,2\eta}|x).$$

This proves that for n large enough

$$\sup_{t\in U_{T,h}(K_{x,\eta}|x)} \sup_{x'\in B(x,h)} (|b_Y| \vee |b_C|)(t|x') \leq \sup_{t\in U_T(K_{x,2\eta}|x)} \sup_{x'\in B(x,h)} (|b_Y| \vee |b_C|)(t|x').$$

Letting $\eta > 0$ be so small that condition (12) holds with η replaced by 2η and using this Hölder condition along with (11) we deduce that

$$\begin{aligned} & \sup_{t\in U_{T,h}(K_{x,\eta}|x)} \sup_{x'\in B(x,h)} (|b_Y| \vee |b_C|)(t|x') \\ = \; & \mathrm{O}\left(h^\alpha \log n_x \vee \sup_{t\in U_T(K_{x,2\eta}|x)} |b_Y(t|x)| \vee \sup_{t\in U_T(K_{x,2\eta}|x)} |b_C(t|x)| \right). \end{aligned}$$

Finally, Potter bounds for the regularly varying functions $|b_Y(\cdot|x)|$ and $|b_C(\cdot|x)|$ (see Bingham *et al.* [6], p.25), both having negative regular variation indices, entail

$$\limsup_{n\to\infty} \sup_{t\in U_T(K_{x,2\eta}|x)} \frac{|b_Y(t|x)|}{|b_Y(U_T(n_x/k_x|x)|x)|} \vee \frac{|b_C(t|x)|}{|b_C(U_T(n_x/k_x|x)|x)|} < \infty$$

which yields (27) and the result in this case.

We now turn to case 2. We remark that

$$\begin{aligned} \overline{F}_C(t|x')f_Y(t|x') + \frac{1}{\gamma_Y(x)}G(t|x') &= r_Y(t,x,x')G(t|x') \\ \text{and} \;\; \overline{F}_Y(t|x')f_C(t|x') + \frac{1}{\gamma_C(x)}G(t|x') &= r_C(t,x,x')G(t|x') \end{aligned}$$

with

$$G(t|x') = \begin{cases} \dfrac{L_{\overline{F}_Y}((U_T(\infty|x') - t)^{-1}|x')L_{\overline{F}_C}((U_T(\infty|x') - t)^{-1}|x')}{(U_T(\infty|x') - t)^{1/\gamma_Y(x')+1/\gamma_C(x')+1}} & \text{if} \;\; 0 < t < U_T(\infty|x') \\ 0 & \text{otherwise} \end{cases}$$

and

$$\begin{aligned} r_Y(t,x,x') &= \frac{1}{\gamma_Y(x)} - \frac{1}{\gamma_Y(x')} - b_Y((U_T(\infty|x') - t)^{-1}|x'), \\ r_C(t,x,x') &= \frac{1}{\gamma_C(x)} - \frac{1}{\gamma_C(x')} - b_C((U_T(\infty|x') - t)^{-1}|x'). \end{aligned}$$

A particular consequence of this is, according to Lemma 2:

$$p_h(t|x) = \frac{-\dfrac{1}{\gamma_Y(x)} + \dfrac{\mathbb{E}(r_Y(t,x,X)G(t|X)\,|\,X \in B(x,h))}{\mathbb{E}(G(t|X)\,|\,X \in B(x,h))}}{-\dfrac{1}{\gamma_Y(x)} - \dfrac{1}{\gamma_C(x)} + \dfrac{\mathbb{E}([r_Y(t,x,X)+r_C(t,x,X)]G(t|X)\,|\,X \in B(x,h))}{\mathbb{E}(G(t|X)\,|\,X \in B(x,h))}}.$$

Define $I_{x,x',\eta} = [U_{T,h}(n_x/(1+\eta)k_x|x), U_T(\infty|x'))$. We shall now prove that

$$\sup_{x' \in B(x,h)} \sup_{t \in I_{x,x',\eta}} (|b_Y| \vee |b_C|)((U_T(\infty|x') - t)^{-1}|x') = \mathrm{O}(h^\alpha \log n_x \vee \delta_n) \to 0 \tag{28}$$

as $n \to \infty$. The fact that $G(\cdot|X)$ is nonnegative shall then yield

$$\begin{aligned}
& \omega\left(p \circ U_T, \frac{n_x}{(1+\eta)k_x}, n_x^{1+\eta}, x, h\right) \\
= \;& \mathrm{O}\left(h^\alpha \vee \sup_{x' \in B(x,h)} \sup_{t \in I_{x,x',\eta}} (|b_Y| \vee |b_C|)((U_T(\infty|x') - t)^{-1}|x')\right) \\
= \;& \mathrm{O}\left(h^\alpha \log n_x \vee \delta_n\right)
\end{aligned}$$

which is what we want to prove. To this aim, remark that one has (see Lemma 1.2.9 in [19], p.22)

$$\frac{U_T(n_x/k_x|x)}{a_T(n_x/k_x|x)} = -\frac{U_T(\infty|x)}{\gamma_T(x)}[U_T(\infty|x) - U_T(n_x/k_x|x)]^{-1}(1+\mathrm{o}(1)) \;\text{ as }\; n \to \infty.$$

Using (9), it is a consequence of the mean value theorem that

$$\sup_{z \in K_{x,\eta}} \left| \frac{U_{T,h}(z|x)}{U_T(z|x)} - 1 \right| = \mathrm{o}(U_T(\infty|x) - U_T(n_x/k_x|x)) = \mathrm{o}\left(\left[\frac{n_x}{k_x}\right]^{\gamma_T(x)} L_{U_T}(n_x/k_x|x)\right)$$

as $n \to \infty$. Especially, (7) and (8) entail that

$$\sup_{x' \in B(x,h)} \frac{U_T(\infty|x') - U_{T,h}(n_x/(1+\eta)k_x|x)}{U_T(\infty|x) - U_T(n_x/k_x|x)} \to (1+\eta)^{-\gamma_T(x)} \;\text{ as }\; n \to \infty.$$

The interval $I_{x,x',\eta}$ is therefore well-defined for all n large enough and $x' \in B(x,h)$, and there exists some constant $\eta' > 0$ such that for n large enough

$$\forall x' \in B(x,h),\; t \in I_{x,x',\eta} \Rightarrow (U_T(\infty|x') - t)^{-1} \in J_{x,\eta'} = \left[\frac{1-\eta'}{U_T(\infty|x) - U_T(n_x/k_x|x)}, \infty\right).$$

This proves that for n large enough

$$\sup_{x' \in B(x,h)} \sup_{t \in I_{x,x',\eta}} (|b_Y| \vee |b_C|)((U_T(\infty|x') - t)^{-1}|x') \le \sup_{z \in J_{x,\eta'}} \sup_{x' \in B(x,h)} (|b_Y| \vee |b_C|)(z|x')|.$$

Conditions (11) and (14) then entail

$$\begin{aligned}
& \sup_{x' \in B(x,h)} \sup_{t \in I_{x,x',\eta}} (|b_Y| \vee |b_C|)((U_T(\infty|x') - t)^{-1}|x') \\
= \;& \mathrm{O}\left(h^\alpha \log n_x \vee \sup_{z \in J_{x,\eta'}} |b_Y(z|x)| \vee \sup_{z \in J_{x,\eta'}} |b_C(z|x)|\right).
\end{aligned}$$

We conclude by using Potter bounds for the regularly varying functions $|b_Y(\cdot|x)|$ and $|b_C(\cdot|x)|$ to get

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{t\in J_{x,\eta'}} \frac{|b_Y(t|x)|}{|b_Y((U_T(\infty|x) - U_T(n_x/k_x|x))^{-1}|x)|} &<& \infty \\
\text{and} \quad &\limsup_{n\to\infty} \sup_{t\in J_{x,\eta'}} \frac{|b_C(t|x)|}{|b_C((U_T(\infty|x) - U_T(n_x/k_x|x))^{-1}|x)|} &<& \infty
\end{aligned}$$

of which (28) is a direct consequence. ■

Acknowledgements

The author acknowledges both the editor and an anonymous referee for their helpful comments which led to significant enhancements of this article.

References

[1] Balslev, U., Monforte, A.D., Stergiou, G., Antunes, F., Mulcahy, F., Pehrson, P.O., Phillips, A., Pedersen, C., Lundgren, J.D. (1997). Influence of age on rates of new AIDS-defining diseases and survival in 6546 AIDS patients, *Scandinavian Journal of Infectious Diseases* **29**(4): 337–343.

[2] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004). *Statistics of Extremes*, John Wiley and Sons.

[3] Beirlant, J., Guillou, A., Dierckx, G., Fils-Villetard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring, *Extremes* **10**: 151–174.

[4] Beirlant, J., Guillou, A., Toulemonde, G. (2010). Peaks-Over-Threshold modeling under random censoring, *Communications in Statistics – Theory and Methods* **39**(7): 1158–1179.

[5] Beran, R. (1981). Nonparametric regression with randomly censored data, technical report, University of California at Berkeley.

[6] Bingham, N.H., Goldie, C.M., Teugels, J.L. (1987). *Regular Variation*, Cambridge, U.K.: Cambridge University Press.

[7] Daouia, A., Gardes, L., Girard, S., Lekina, A. (2011). Kernel estimators of extreme level curves, *Test* **20**(2): 311–333.

[8] Daouia, A., Gardes, L., Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli* **19**(5B): 2557–2589.

[9] Darby, S.C., Ewart, D.W., Giangrande, P.L., Spooner, R.J., Rizza, C.R. (1996). Importance of age at infection with HIV-1 for survival and development of AIDS in UK haemophilia population. UK Haemophilia Centre Directors' Organisation, *The Lancet* **347**(9015): 1573–1579.

[10] Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics* **17**(4): 1833–1855.

[11] Einmahl, J.H.J., Fils-Villetard, A., Guillou, A. (2008). Statistics of extremes under random censoring, *Bernoulli* **14**(1): 207–227.

[12] Fisher, R.A., Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society* **24**: 180–190.

[13] Gardes, L., Girard, S. (2012). Functional kernel estimators of large conditional quantiles, *Electronic Journal of Statistics* **6**: 1715–1744.

[14] Gardes, L., Stupfler, G. (2014). Estimation of the conditional tail-index using a smoothed local Hill estimator, *Extremes* **17**(1): 45–75.

[15] Gardes, L., Stupfler, G. (2015). Estimating extreme quantiles under random truncation, *TEST* **24**(2): 207–227.

[16] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics* **44**: 423–453.

[17] Goegebeur, Y., Guillou, A., Schorgen, A. (2014). Nonparametric regression estimation of conditional tails – the random covariate case, *Statistics* **48**(4): 732–755.

[18] Goegebeur, Y., Guillou, A., Stupfler, A. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **51**(3): 1190–1213.

[19] de Haan, L., Ferreira, A. (2006). *Extreme value theory: An introduction.* Springer, New York.

[20] Hall, P. (1982). On some simple estimates of an exponent of regular variation, *Journal of the Royal Statistical Society: Series B* **44**: 37–42.

[21] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics* **3**: 1163–1174.

[22] Luo, K., Law, M., Kaldor, J.M., McDonald, A.M., Cooper, D.A. (1995). The role of initial AIDS-defining illness in survival following AIDS, *AIDS* **9**(1): 57–63.

[23] Ndao, P., Diop, A., Dupuy, J.-F. (2014). Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring, *Computational Statistics and Data Analysis* **79**: 63–79.

[24] Ripley, B.D., Solomon, P.J. (1994). A note on Australian AIDS survival, University of Adelaide Department of Statistics Research Report 94/3.

[25] Sayah, A., Yahia, D., Brahimi, B. (2014). On robust tail index estimation under random censorship, *Afrika Statistika* **9**(1): 671–683.

[26] Stupfler, G. (2013). A moment estimator for the conditional extreme-value index, *Electronic Journal of Statistics* **7**: 2298–2343.

[27] Venables, W.N., Ripley, B.D. (2002). *Modern Applied Statistics with S*, Springer, fourth edition.

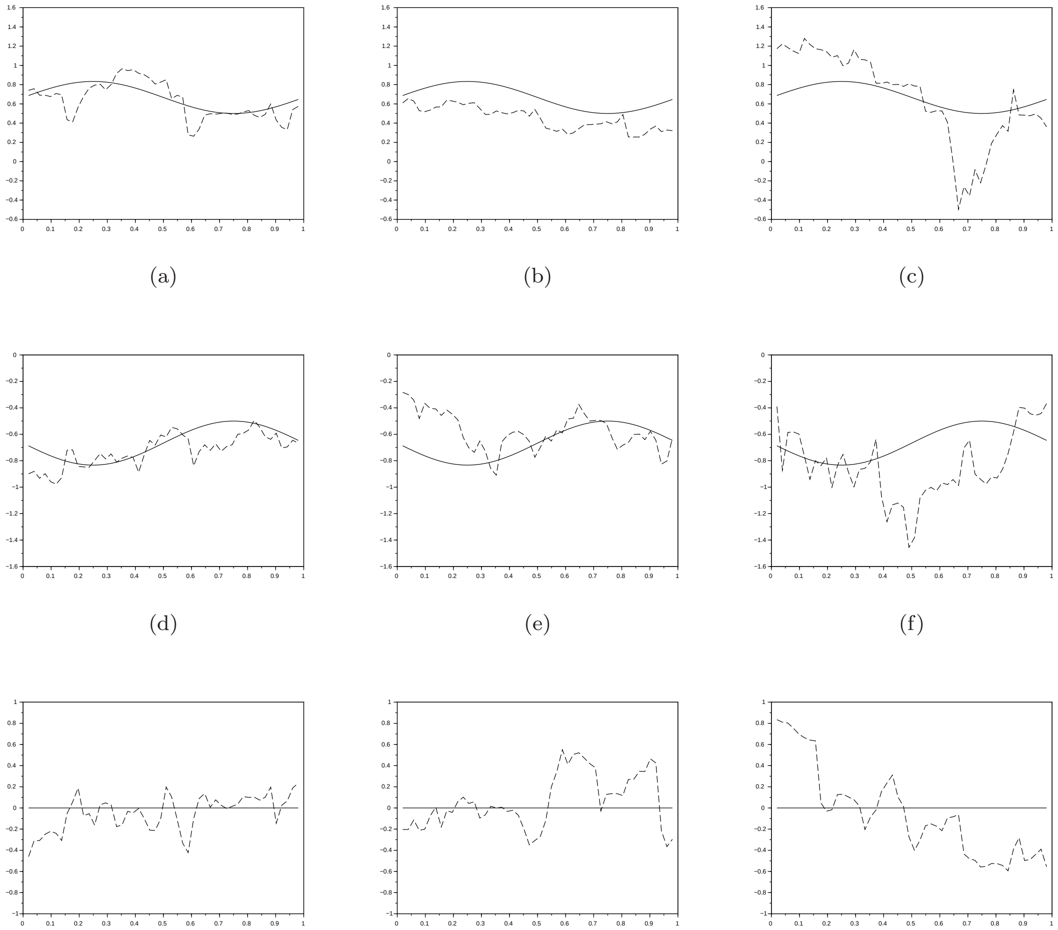
[28] Wang, H., Tsai, C.L. (2009). Tail index regression, *Journal of the American Statistical Association* **104**(487): 1233–1240.

[29] Worms, J., Worms, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions, *Extremes* **17**(2): 337–358.

Censored cases	
Situation	Moment estimator $\widehat{\gamma}_Y$
Fréchet-Fréchet model	
$\rho = -0.5$	0.177 $[0.0138, 0.550]$
$\rho = -1$	0.0639 $[0.0139, 0.170]$
$\rho = -1.5$	0.0491 $[0.00563, 0.146]$
Weibull-Weibull model	
$c = 0.1$	0.0451 $[0.00956, 0.138]$
$c = 0.2$	0.0505 $[0.0146, 0.165]$
$c = 0.3$	0.0494 $[0.0125, 0.137]$
Gumbel-Gumbel model	0.0840 $[0.0172, 0.334]$

Uncensored cases	
Situation	Moment estimator $\widehat{\gamma}_Y$
Gumbel-Fréchet model	0.0352 $[0.00476, 0.102]$
Weibull-Fréchet model	0.0375 $[0.00587, 0.144]$
Weibull-Gumbel model	0.0364 $[0.00750, 0.0997]$

Table 1: MSEs associated to the estimator $\widehat{\gamma}_Y$ in all cases. Between brackets: minimal and maximal mean squared error recorded.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 1: Censored cases: the true function γ_Y (solid line) and its estimator $\widehat{\gamma}_Y$ (dashed line). Top row: Fréchet-Fréchet model, case $\rho = -1$. Middle row: Weibull-Weibull model, case $c = 0.1$. Bottom row: Gumbel-Gumbel model. Left: case corresponding to the 5% quantile of the MSE. Middle: case corresponding to the median of the MSE. Right: case corresponding to the 95% quantile of the MSE.

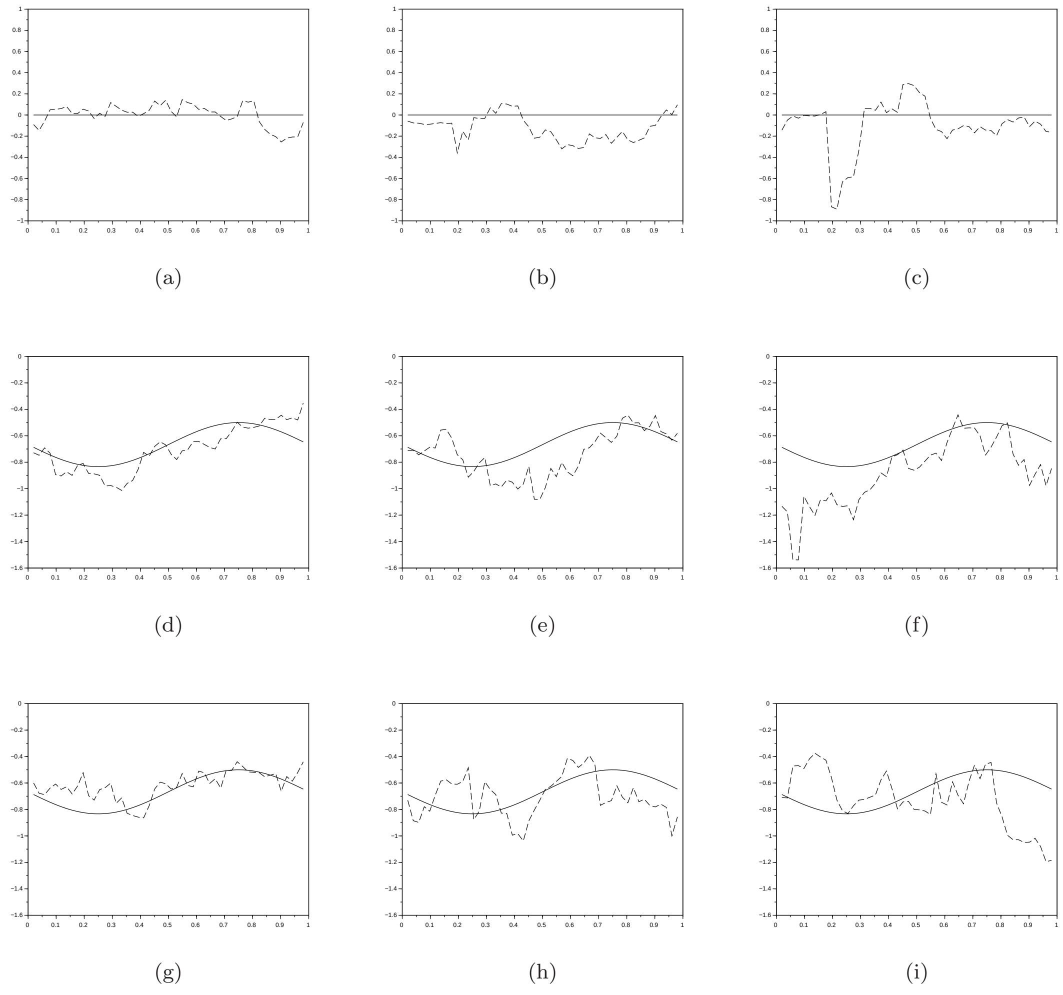


Figure 2: Uncensored cases: the true function γ_Y (solid line) and its estimator $\widehat{\gamma}_Y$ (dashed line). Top row: Gumbel-Fréchet model. Middle row: Weibull-Fréchet model. Bottom row: Weibull-Gumbel model. Left: case corresponding to the 5% quantile of the MSE. Middle: case corresponding to the median of the MSE. Right: case corresponding to the 95% quantile of the MSE.

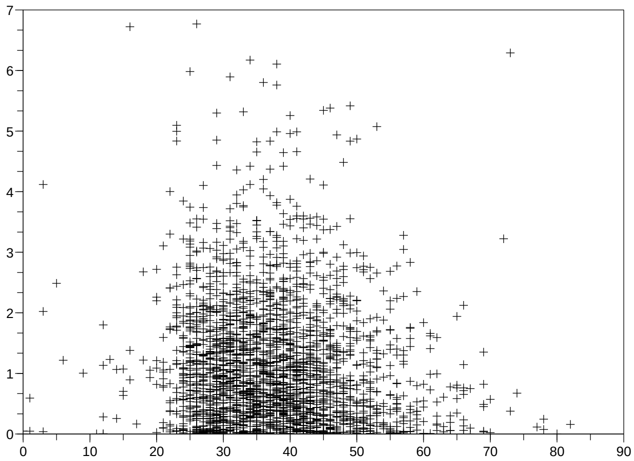


Figure 3: Scatterplot of the AIDS data: $x-$axis: age of the patient at the time of diagnosis, $y-$axis: survival time (in years).

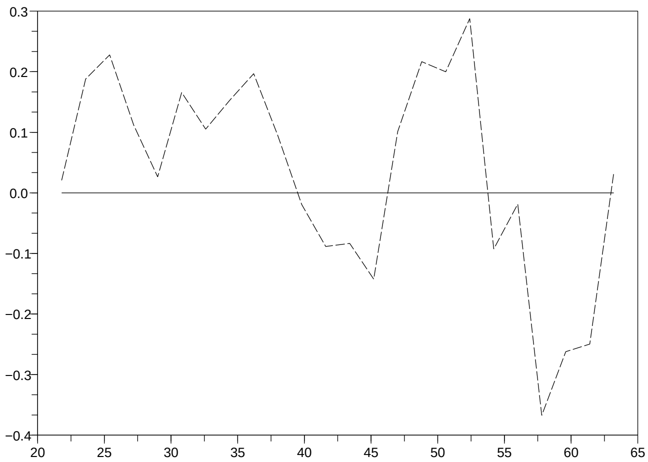


Figure 4: AIDS data: estimator $\widehat{\gamma}_Y$. $x-$axis: age of the patient at the time of diagnosis.

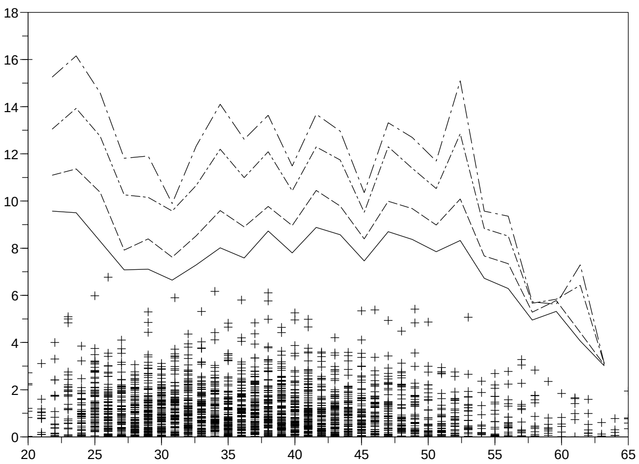


Figure 5: AIDS data: estimation of the conditional extreme quantile of the survival time. Full line: level $\varepsilon = 0.01$, dashed line: level $\varepsilon = 0.005$, dashed-dotted line: $\varepsilon = 0.002$, dotted line: level $\varepsilon = 0.001$. $x-$axis: age of the patient at the time of diagnosis, $y-$axis: survival time (in years).